PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

6869

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification [5] : C12N 15/12, C12Q 1/68, 1/02 G01N 33/53	A1	(11) International Publication Number: WO 91/05044
		(43) International Publication Date: 18 April 1991 (18.04.91)

(21) International Application Number: PCT/GB90/01481

(22) International Filing Date: 27 September 1990 (27.09.90)

(30) Priority data:
8921791.3 27 September 1989 (27.09.89) GB

(71) Applicant *(for all designated States except US)*: IMPERIAL CANCER RESEARCH TECHNOLOGY LTD. [GB/GB]; Sardinia House, Sardinia Street, London WC2A 3NL (GB).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)* : BATES, Gillian [GB/GB]; LEHRACH, Hans, Rudolf [AT/GB]; Imperial Cancer Research Fund, P.O. Box 123, Lincolns Inn Fields, London WC2A 3PX (GB). GUSELLA, James, F. [CA/US]; 7 Woodstock Drive, Framingham, MA 01701 (US). MACDONALD, Marcy, E. [CA/US]; 183 Larch Road, Cambridge, MA 02138 (US).

(74) Agents: CRESSWELL, Thomas, Anthony et al.; J.A. Kemp & Co., 14 South Square, Gray's Inn, London WC1R 5LX (GB).

(81) Designated States: AT (European patent), BE (European patent), CH (European patent), DE (European patent)*, DK (European patent), ES (European patent), FR (European patent), GB (European patent), IT (European patent), JP, LU (European patent), NL (European patent), SE (European patent), US.

Published
With international search report.

(54) Title: A CULTURE OF *SACCHAROMYCES CERVISIÆ* CONTAINING THE GENE RESPONSIBLE FOR HUNTINGTON'S DISEASE AND USES THEREOF

(57) Abstract

The present invention relates to nucleic acids useful in the detection and/or treatment of Huntington's disease. DNA has been isolated from the p16.3 band of human chromosome 4 and put into *Saccharomyces cerrisiæ* clone 488 BT (NCYC2336). The invention also provides nucleic acid fragments capable of hybridising with said clone, polypeptides produced by it, and monoclonal antibodies against the polypeptides.

DESIGNATIONS OF "DE"

Until further notice, any designation of "DE" in any international application whose international filing date is prior to October 3, 1990, shall have effect in the territory of the Federal Republic of Germany with the exception of the territory of the former German Democratic Republic.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	**ES**	Spain	**MC**	Monaco
AU	Australia	**FI**	Finland	**MG**	Madagascar
BB	Barbados	**FR**	France	**ML**	Mali
BE	Belgium	**GA**	Gabon	**MR**	Mauritania
BF	Burkina Fasso	**GB**	United Kingdom	**MW**	Malawi
BG	Bulgaria	**GR**	Greece	**NL**	Netherlands
BJ	Benin	**HU**	Hungary	**NO**	Norway
BR	Brazil	**IT**	Italy	**PL**	Poland
CA	Canada	**JP**	Japan	**RO**	Romania
CF	Central African Republic	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CG	Congo			**SE**	Sweden
CH	Switzerland	**KR**	Republic of Korea	**SN**	Senegal
CM	Cameroon	**LI**	Liechtenstein	**SU**	Soviet Union
DE	Germany	**LK**	Sri Lanka	**TD**	Chad
DK	Denmark	**LU**	Luxembourg	**TG**	Togo
				US	United States of America

A CULTURE OF SACCHAROMYCES CERVISIAE CONTAINING THE GENE RESPONSIBLE FOR HUNTINGTON'S DISEASE AND USES THERE OF.

The present invention relates to nucleic acids useful in the detection and/or treatment of Huntington's disease and to materials derived from the nucleic acids.

Huntington's disease (HD) is a progressive neurological disorder that typically produces emotional disturbances, lack of motor coordination and intellectual deterioration. The onset of symptoms generally occurs in mid-life, followed by progression of the disease to death over 15-20 years (Martin and Gusella 1986). It is characterised pathologically by an extensive loss of specific neuronal classes, occuring primarily within the caudate nucleus and putamen with the relative sparing of neurons in the rest of the brain. The biochemical defect underlying the mechanism of this cell death is unknown and there is no existing treatment that will arrest the course of the disease.

The inheritance of HD is well documented; it is caused by an autosomal dominant gene with complete penetrance and low mutation rate (Martin and Gusella 1986). Genetic linkage to the anonymous DNA marker D4S10 localised the mutation to the short arm of chromosome 4 (Gusella et al. 1983) and multipoint analysis has indicated that the gene is distal to this marker within the most telomeric subband, 4p16.3 (Gilliam et al. 1987a). The discovery of linked markers has allowed the development of a test for

both the presymptomatic and prenatal id ntification of probable gen carriers (Meisson et al. 1988). It has also led to the identification of individuals with a high probability of being homozygous for HD and, interestingly, the progression of HD in these individuals is indistinquisable from that in typical HD heterozygotes (Wexler et al. 1987; Myers et al. 1989). Recently, an intense effort has been employed to isolate closer and especially flanking markers to precisely define the position of the gene (Gilliam et al. 1987b; MacDonald et al. 1987; Pohl et al. 1988; Richards et al. 1988; Smith et al. 1988; Wasmuth et al. 1988; Whaley et al. 1988; Youngman et al. 1988; MacDonald et al. 1989b).

A long range restriction map has been constructed with more than 20 independently derived DNA probes that map distal to D4S10. The map consists of three as yet unlinked segments that together cover 5 million base pairs (Mb) and extends minimally 4 Mb distal from D4S10 (Bucan et al. 1989) to a position characterised either as a cluster of well cut restriction sites or the telomere of 4p. Order and orientation of the segments has been determined by genetic linkage studies (Richards et al. 1988; Wasmuth et al. 1988; Whaley et al. 1988; Youngman et al. 1989; MacDonald et al. 1989a) and by somatic cell genetics (MacDonald et al. 1987; Smith et al. 1988). Analysis of recombination events in Huntington's disease families (MacDonald et al. 1989a; Robbins et al. 1989) suggests that the most likely position for the HD gene is telomeric to

the most distal published marker D4S90 (defin d by probe D5) situat d 300 kb from the end of the map (Bucan et al. 1989). In spite of the intense analysis, a definitive flanking marker has not yet been identified.

The inventors have isolated DNA from the p16.3 band of human chromosomes 4 which contains the gene responsible for and characteristic of Huntington's disease (HD). A culture of Saccharomyces cerevisiae containing a yeast artificial chromosome (YAC clone) designated Y88BT containing this DNA was deposited on 26th September 1989 under Budapest Treaty Conditions in the National Collection of Yeast Cultures, Colney Lane, Norwich, NR4 7AU, England under the Accession No _____.

The present invention provides, in one aspect, YAC- clone Y88BT (Accession no. _____).

In another aspect the invention provides nucleic acid probes capable of hybridising with YAC-clone Y88BT. The probes may be single or double stranded, RNA or DNA. Preferably they will be labelled with conventional labels such as radio-labels, fluorescent labels or enzyme labels. The probes will be capable of hybridising with Y88BT under low stringency conditions. Preferably the probes will hybridise under high stringency conditions.

The present invention also provides nucleic acid fragments capable of hybridising with Y88BT under low or high stringency conditions, containing a coding sequence corresponding to the wild-type or mutant gene which, in mutant form is responsible for HD.

Such coding sequences may encode the whole or a part of the polypeptide encoded by the HD gene or the wild-type gene of which the HD gene is a mutant. They may contain introns with suitable splicing sites and ligation signal sequences. Preferably the nucleic acid fragments also contain transcription and translation initiation and termination sequences in correct reading frame register with the coding sequences. They may further comprise transcription enhancer and promoter sequences operationally linked to the coding sequences or other sequences responsible for the phenotype of the mutation. The nucleic acid fragments may be in the form of inserts in cloning or expression vectors such as cosmids, plasmids, yeast artificial chromosomes and viral genomic nucleic acids.

The present invention also provides virus particles containing nucleic acid fragments as hereinbefore described in the genomic nucleic acid and cells transfected with cloning or expression vectors or infected with such virus particles containing a nucleic acid fragment as hereinbefore described.

The present invention further provides processes for expressing polypeptides encoded by nucleic acid fragments as hereinbefore defined comprising culturing cells transfected with cloning or expression vectors. Such culturing and expression may be conducted *in vitro* in tissue or cell culture or *in vivo* in ascites or in transgenic animals. Preferably polypeptides expressed by these methods may be recovered from the culture medium,

ascites fluid or a body fluid of transgenic animals or from the cells expressing the polypeptides by disrupting the cells and fractionating the cell debris. Transgenic animals harbouring the wild type or mutant gene which in mutant form is responsible for HD form a further aspect of the present invention.

Polypeptides encoded by the nucleic acids as hereinbefore described form a further aspect of the invention. They may be useful for immunisation or therapy of HD and in diagnostic test methods for identifying carriers of the HD gene.

The polypeptides may be immbolised or labelled for instance for use in such diagnostic tests.

The invention further provides antibodies, which may be, for instance, polyclonal or monoclonal antibodies, and fragments of such antibodies, for instance Fab or $F(ab')_2$ fragments, raised against or specific for the polypeptides described above, and processes for production of the antibodies and fragments as described below.

Such antibodies and fragments may be produced by conventional immunisation techniques using the polypeptides described above. Polyclonal antibodies are recovered from body fluids of animals immunised with the polypeptides. Alternatively the animals may be inoculated with cells containing nucleic acids as hereinbefore defined or infected with viruses containing such nucleic acids in a manner such that the cells or viruses express the polypeptides and stimulate an immune reaction. Polyclonal

antibodies may be recovered from the body fluids of animals so inoculated or infected.

Monoclonal antibodies are produced by removing antibody-secreting cells from animals inoculated, infected or immunised as described above, and immortalising the cells by fusion with immortals cells or other known techniques such as infection with Epstein-Barr virus, culturing the immortal cells and recovering the antibodies secreted thereby.

The present invention also provides antibody-secreting cells and immortalised such cells or the decendants thereof which produce antibodies raised against or specific for polypeptides as described above.

The antibodies and fragments thereof are useful in passive immunisation against the peptides described above and in diagnostic and therapeutic processes.

For diagnostic uses the antibodies or fragments may be immobilised or labelled with labels such as radio-labels, fluorescent labels or enzyme labels. For therapeutic uses the antibodies or fragments may be used as targetting entities for therapeutic or cytotoxic agents, or in passive immunisation.

For therapeutic uses the nucleic acids, probes, vectors, cells, viruses, polypeptides, antibodies or fragments thereof are preferably presented as unit- or multi-dose formulations comprising a pharmaceutically acceptable diluent or carrier, for instance as tablets, capsules, solutions, suspensions or dry powders or

concentrates suitable for use in producing s luti ns or suspensions, for administration orally or parenterally (in which cas the solutions or suspensions will be sterile and pyrogen-free) for instance intramuscularly, subcutaneously or intravenously as an injection or, in the case of intravenous administration, as a continuous infusion. Dosages and dosage regimes will be determined according to the activity and toxicity of the agent administered, the objective of administration and the age, size, weight, health and nutritional status of the recipient taking into account absorption and clearance rates and tissue distribution patterns for the agent administered.

Such formulations form a further aspect of the invention.

The invention further provides nucleic acids and probes and vectors, cells and viruses containing nucleic acids as hereinbefore defined, polypeptides and antibodies and fragments thereof raised against or specific for polypeptides as hereinbefore described and antibody-secreting cells for use in the treatment of the human or animal body by surgery or therapy or in diagnostic methods practised on the human or animal body and their use in preparation of medicaments for use in such treatment or diagnostic methods as well as methods for treatment or diagnosis comprising administering a non-toxic effective amount of a nucleic acid, probe, vector, cell, virus, polypeptide or antibody or fragment thereof to a human or non-human animal in need thereof.

The invention further provides diagnostic tests practised n a sampl of tissue, cells or body fluids of individuals carrying or suspected to carry the HD gene comprising contacting the sample with a nucleic acid, probe, vector, cell, virus, polypeptide, antibody or fragment thereof, and kits for use in such tests comprising at least the nucleic acid, probe, vector, cell, virus, polypeptide, antibody or fragment thereof and optionally further comprising, separately packaged, buffers, labelling reagents, reagents for detecting labels, controls and/or standards.

The invention will be further explained and illustrated by the following description of the production and characterisation of Y88BT. The novel materials and processes described below represent particular embodiments of the invention.

We have set out to delineate and clone the telomeric candidate region and describe here the use of Saccharomyces cerevisiae as a direct route to selectively clone the telomere of the short arm of chromosome 4. This clone would minimally set physical limits to the position of the gene and provide a definitive flanking marker to the mutation. Whilst telomere cloning in E. coli has been reported (Blackburn and Challoner 1984; Ponzi et al. 1985), YAC vectors, allowing the replication of mammalian DNA fragments as artificial chromosomes in yeast (Burke et al. 1987), have provided the opportunity of

a simpler strategy. Such vectors contain a yeast centromere (CEN4), a replicating sequence (ARS1), selectable markers (URA3 and TRP1), and two cassettes of the telomeric repeat from Tetrahymena (TEL). The vector telomeres are maintained by the addition of Saccharomyces telomeric repeats by the yeast telomerase (Shampay et al. 1984), a process that is template independent (Greider and Blackburn 1985). Telomeric repeats, which consist of tandem arrays of short (6-8bp) G-T rich sequences, are highly conserved, seemingly across all eukaryotic species (Blackburn 1984; Moyzis et al. 1988). The sequencing of the human telomeric repeat has recently shown it to be identical to the Trypanosoma repeat and to differ by only one base with that from Tetrahymena (Moyzis et al. 1988). It seemed likely, therefore, that a human telomere would be functional in yeast. We describe here the construction of a vector containing ARS1, CEN4, URA3 and one cassette of the Tetrahymena telomeric repeat (TEL), to provide a system for the selection of mammalian telomeres (YAC-t1). Similar approaches have recently been described (Brown 1989; Cross et al. 1989; Cheng et al. 1989).

We have isolated overlapping cosmids, 2R88 and B31 (defining locus D4S142) which on further analysis of the distal region of 4p16.3 by pulsed field gel electrophoresis (PFGE) were found to map approximately 200 kb distal to D4S90. An MspI restriction fragment length polymorphism (RFLP) detected by p88-18 (a subclone from 2R88) suggests that the HD mutation may be yet distal to this locus. Several lines of evidence suggested that the end of the pulsed field gel map is in fact the end of the chromosome. The addition of a rare cutter polylinker to the YAC-t1 vector (YAC-t2) allowed the isolation of the most distal 120 kb BssHII fragment of 4p16.3 by the construction of a

BssHII-telomere library and its subsequent screening with a probe from 2R88. The strategy of constructing the library from an individual homozygous for the HD gene guarantees that the clone described here derives from an HD chromosome.

Materials and Methods

DNA and Cell lines

Lymphoblastoid cell lines were as follows: GM1416B (karyotype 48 XXXX) (NIGMS, Human Genetic Cell Repository, Camden NJ); GUSHMI and GUSHM3 are from individuals from the Venezuela pedigree likely to be homozygous for *HD* (Wexler et al. 1987); GUS641 and GUS115 are from individuals from the Venezuela pedigree that do not carry the HD mutation. Somatic cell hybrids included: HHW693, a human-hamster hybrid containing human 4p15.1-4pter translocated onto a fragment of the short arm of human chromosome 5 (Wasmuth et al. 1986); HHW842, a human-hamster hybrid containing human chromosome 5 and human chromosome 4 with an interstitial deletion of 4p14-4p16.3, retaining the terminal portion of 4p16.3 (Smith et al. 1988); HHW847, a human-hamster hybrid containing several human chromosomes, including human chromosome 5, and a t(4:21) chromosome in which 4p16.2-4pter is absent (Smith et al. 1988).

DNA Preparation, Digestion, Fractionation, Transfer and Hybridization

DNA for conventional Southern blot analysis was extracted from peripheral blood leukocytes and lymphoblastoid cell lines by standard techniques. Isolation of high molecular weight DNA in agarose blocks from lymphoblastoid cell lines for both PFGE ($7.5x10^5$ cells/block) and cloning into YAC vectors ($3x10^6$ cells/block) was as previously published (Herrmann et al. 1987). High molecular weight DNA from blood was prepared by the selective lysis of erythrocytes (Herrmann and Frischauf 1987) with modifications as

described by Bucan et al. (1989). Chromosomes from Saccharomyces cerevisiae, for use as molecular weight standards on pulsed field gels (strain YP148) and for the analysis of YAC recombinants, were prepared as described by Carle and Olson (1985). Lambda multimers were purchased from FMC Bioproducts.

Restriction enzymes were from NEB or BRL and digests were performed according to manufacturers recommendations. Restriction of DNA for, and fractionation of DNA by, PFGE was essentially as described in Herrmann et al. (1987). PFGE was performed by contour clamped homogeneous electric fields (CHEF) in an apparatus similar to that described by Chu et al. (1986) constructed at the EMBL, Heidelberg. Electrophoresis was at 5 V/cm in 0.25 x TBE at 14°C, with specific gel conditions and pulse times as described in the text and figure legends. DNA was transferred in alkali for 2 hrs onto Hybond-N+ (Amersham) from conventional agarose gels and for 48 hrs onto Hybond-N (Amersham) from pulsed field gels. Probes were labeled in agarose to high specific activity by random oligonucleotide priming (Feinberg and Vogelstein 1984). Hybridizations were performed in 50% formamide at 42°C (Monaco et al. 1985) and filters were washed as described by Church and Gilbert (1984). Probes containing low copy repeat sequences were prehybridized with cold, sonicated total human DNA (Sealey et al. 1985). The oligomer $(CCCTAA)_4$ was labeled by kinasing and hybridized in 0.5M Na_2HPO_4 pH 7.2, 7% SDS, 1 mM EDTA at 42°C for 20 hrs. The filters were rinsed three times in 3 x SSC, 0.1% SDS (1 x SSC is 0.15 M NaCl, 0.015 M trisodium citrate pH 7.0) at RT and washed twice for 2 min in 3 x SSC, 0.1% SDS at 55°C.

Restriction Mapping of cosmids

Restriction mapping of cosmids was according to Rackwitz et al. (1985), using a computer program package for restriction map analysis and manipulation described by Zehetner and Lehrach (1986).

Construction of telomere cloning vectors YAC-t1 and YAC-t2

The construction of the telomere cloning vectors is outlined in Figure 5. YCp50 (Hieter et al. 1985) is a pBR322 based plasmid which contains ARSI (yeast origin of replication), CEN4 (yeast centromere) and URA3 (selectable marker) whilst retaining the plasmid origin of replication, amp and tet genes. EcoRI, BamHI and SalI are among its unique restriction sites. The 0.7 kb BamHI/XhoI fragment from YAC4 (Burke et al. 1987), that contains the Tetrahymena telomere repeats (TEL), was isolated from a LMP agarose (BRL) gel by treatment with agarase (Burmeister and Lehrach 1989) and ligated into the BamHI/SalI sites of YCp50 to create YAC-t1 (8.38 kb). The correct construct was indicated by tet sensitivity and by the destruction of the SalI site and retention of a BamHI site. Cleavage with BamHI/EcoR1 generates a linear telomere cloning vector with TEL sequences at one end and an EcoRI cloning site at the other. Replacement of the BamHI/EcoRI fragment with a rare cutter polylinker containing NotI, SacII, SalI, MluI, ClaI and SnaBI sites, and flanked by BamHI and EcoRI complementary ends, led to construction of YAC-t2 (8.05 kb). This polylinker was analogous to one previously described by Marchuk and Collins (1988) in the YAC-RC vector, and in addition to the unique restriction sites listed above, will accept DNA generated by BssHII, EagI, NaeI, NarI, NruI, SmaI and XhoI digests due to their ligation compatability with sites in the polylinker.

Preparation of the YAC telomere library

YAC-t2 was cleaved with MluI (NEB) and BamHI (BRL) and phosphatased (Boehringer-Mannheim). High molecular weight genomic DNA was prepared from GUSHM1 in agarose blocks at $3x10^6$ cells/80 ul in 0.6% LMP agarose (BRL). This was digested to completion with BssHII (NEB, 20U) for 3 hrs. Digestion was terminated by incubation with Proteinase K (1 mg/ml) (BDH) and 50 mM EDTA pH 8.0 at 37°C for 30 min. The Proteinase K was inactivated with 2 x 30 min incubations in 15 ml 40 ug/ml PMSF (phenylmethylsulphonyl-fluoride; Sigma) in TE (10 mM Tris (pH 8.0) 1 mM EDTA) at 37°C. The blocks were equilibrated with NaCl such that the final salt concentration in the ligation reaction was 50 mM, and melted at 68°C for 5 min. 30 ug of genomic DNA was gently mixed with 60 ug of BamHI/MluI cut and phosphatased YAC-t2 and ligated overnight at 15°C in a total volume of 200 ul in 1 mM ATP, 40 mM Tris-HCl pH 7.6, 10 mM $MgCl_2$, 1 mM DTT with 5 ul T4 DNA ligase (400 U/ul; NEB). Two ligations of 30 ug of BssHII cut GUSHM1 to MluI/BamHI cut and phosphatased YAC-RC were prepared in parallel by the same protocol with the exception that only 30 ug of vector was used.

The YAC-t2 ligation and one of the YAC-RC ligations were melted at 68°C for 5 min, diluted to 300 ul with 3% LMP agarose (BRL) in TE and reset into blocks, which were loaded onto a 0.8% agarose gel in TAE. Much of the non-recombinant vector was electrophoresed out of the ligations for 2 hrs at 5 V/cm. The gel containing the vector was removed and the genomic DNA electrophoresed back into the blocks under the same conditions using an inverted field. The blocks were removed from the gel and equilibrated,

alongside the non-treated YAC-RC ligation, with 50 mM NaCl, 10 mM EDTA and melted at 68°C for 5 min. All ligations were agarased with 100 U agarase (Calbiochem) at 37°C for 3.5 hrs. After 1:1 dilution with 2 M Sorbitol they were frozen on dry ice and stored at -70°C. Ligations were thawed and transformed into AB1380 spheroplasts. The preparation of spheroplasts and the transformation procedure were as described by Burgess and Percival (1987), with the exception that the transformation was performed at 1.2 x 10^9 spheroplasts/ml. 10 ng of YCp50 supercoil DNA was transformed in parallel to provide an estimate of the transformation efficiency. Plates were incubated for 5 days at 30°C in a fan assisted incubator.

Replication and screening of clones

The colonies were replicated in duplicate out of top agar using an aluminium plate with 40,000 machined pins (constructed at the EMBL, Heidelberg) onto ura- plates. After 2 days at 30°C they were lifted onto Hybond-N membranes and lysed overnight on Whatman 3MM soaked in 1 mg/ml zymolyase (FMC Bioproducts) in 1 M Sorbitol, 0.1 M sodium citrate pH 5.8, 10 mM EDTA pH 7.6 and 30 mM 2-mercaptoethanol. Filters were denatured for 10 min on Whatman 3MM soaked in 0.5 M NaOH/1.5 M NaCl and excess denaturant was removed before neutralization by flotation on 1.5 M NaCl/1 M Tris pH 7.4 (neutralization solution) for 2 min. Cell debris was wiped off with Kleenex tissues soaked in 0.1x neutralization solution, with which the filters were then rinsed thoroughly. They were incubated on Whatman 3MM soaked in 200 ug/ml Proteinase K (BDH) in 0.1x neutralization solution for 20 min and then floated on 50 mM Na_2HPO_4 pH 7.2 for 2 min. After air drying, the filters were baked under vacuum for 20 min at 80°C and UV cross linked (Church and Gilbert 1984). Hybridizations were as described above with a probe concentration of 1-2x10^6 cpm/ml.

Results

Characterisation of cosmid clones 2R88 and B31

2R88 and B31 are two of a series of clones (Whaley et al. in preparation) identified by hybridization with total human DNA to a cosmid library constructed from the somatic cell hybrid line HHW693: which contains 4p15.1-4pter translocated onto a fragment of 5p as the only human component on a Chinese hamster cell background (Wasmuth et al. 1986). Figure 1a shows restriction maps of the cosmids, including the positions of sites for the rare cutter enzymes MluI, SalI and BssHII which were used in the further analysis. The positions of unique or low copy fragments that were isolated for use as probes are also indicated.

To determine which of the rare cutter sites identified in the clone are also cleavable in genomic DNA, we hybridized the probes 88BH2.0 and 88E1.8 to Southern blot filters of DNA from the cell line GM1416B, that had been digested either with HindIII only, or with both HindIII and a number of rare cutting enzymes. Figure 1b shows the result of this experiment, indicating cleavage of the rare cutter sites (BssHII, SalI and ClaI) in GM1416B DNA. The ClaI site is not present in 2R88 (since this cosmid is derived from HHW693 DNA).

2R88 is located in the most distal region of 4p16.3, 100 kb from the end of the map

Provisional pulsed field gel mapping data, with the enzymes NotI, MluI and NruI, suggested that 2R88 maps 60-280 kb distal to D5 on a partial 220 kb MluI fragment. In order to position 2R88 precisely, and to determine its

rientation, the region was mapped more extensively with BssHII and SalI using the probe D5, and 88BH4.8, p2R88-1 and 88ES2.3, isolated from 2R88 (Figure 1a). The probes were hybridized to a number of filters containing single and double digests of GM1416B and peripheral leukocyte DNA. Filters were prepared in duplicate with pulse times of 100 sec and 50 sec, which allow a gel resolution of up to 1500 and 600 kb respectively. Figure 2 shows the hybridization of p2R88-1 and D5 to one filter illustrating detection of the same 350 kb NruI fragment in leukocyte DNA and the same partial 300 kb MluI fragment in GM1416B DNA.

Data of this type, summarised in Table I, allowed the construction of a high resolution pulsed field restriction map of the most distal region of 4p16.3 (Figure 3) and localized 2R88 200 kb distal to D5 and 100 kb from the end of the map. However, establishment of the pulsed field map was complicated by the observation of additional bands, identified by cross-hybridization of the probes used to other loci, a phenomenon characteristic of this region of the genome (unpublished observation). Weaker bands, corresponding to fragments detected by cross-hybridizing loci, are indicated by an asterisk in Table 1 and discussed in the table legend.

The end of the map is likely to coincide with the telomere of chromosome 4p

The cluster of restriction sites at the distal end of the map (marked in italics, Figure 3) could represent either a CpG island, containing a well cut site for all enzymes tested, or the 4p telomere. Pulsed field gel analysis is unable to distinguish between these two possibilities, however, several further lines of evidence suggested that this was indeed the telomere. Attempts to identify DNA

beyond these sites by partial digestion were repeatedly unsuccessful. Similarly, we were unable to isolate clones from the potential CpG island by chromosome jumping (Poustka and Lehrach, 1986: Poustka et al. 1987; Poustka and Lehrach 1988) in either MluI or BssHII jumping libraries (Poustka, unpublished), while multiple clones extending to a site proximal to this position were recovered (in preparation). In combination, this evidence was sufficiently compelling to devise a strategy for the isolation of the 4p telomere based on the pulsed field gel analysis presented here..

Cloning of mammalian telomeres in Saccharomyces Cerevisiae

The similarity of the human telomeric repeat to that of Tetrahymena (functional as a telomere in yeast) suggested that a YAC vector carrying only one Tetrahymena telomere would allow the cloning of a human telomere by complementation. To simplify the identification of a clone from 4p, and to allow the construction of clones of sufficient length to be tested for possible biological activity (e.g. in a transgenic mouse system), we were especially interested in a telomere library containing long clones. Such large insert YAC libraries can be constructed by two procedures. One protocol, used by Burke et al. (Burke et al. 1987; Little et al. 1989), relies on the use of very partial digestion with commonly cutting restriction enzymes to generate the large DNA fragments (of the order of hundreds of kilobase pairs) to be cloned. We decided to concentrate on an alternative approach and use rare cutter enzymes, enabling us to take advantage of the information in the long range map of the region.

Choice f rare cutter restriction enzyme for the construction of the telomere library

The majority of the clones recovered from the construction of a rare cutter YAC library, in the absence of prior size selection, will fall within a size range of up to 200 kb (unpublished observation). By pulsed field gel analysis, we have identified a 100 kb BssHII fragment which contains p2R88-1 at its proximal end and extends to the telomere. The preparation of a library from genomic DNA digested with BssHII would therefore generate a clone in a size class for which there is a natural enrichment and allow its identification by hybridization with p2R88-1. Furthermore, hybridization of this probe to DNA from a number of blood samples and cell lines, that had been digested with BssHII and fractionated by PFGE, detected a BssHII polymorphism as shown in Figure 4. Lanes 1 and 2 contain samples from two individuals from the Venezuela pedigree (GUSHM1 and GUSHM3), expected to be homozygous for the HD mutation, and in these DNAs, p2R88-1 detects BssHII fragments of approximately 100 kb and 120 kb. The BssHII fragment of 60 kb previously observed in GM1416B and leukocyte DNA is not present, indicating either complete methylation or absence of the internal site. Therefore, the construction of a BssHII-telomere clone library from one of these individuals would provide an increased chance of recovery of a telomeric clone, and would allow access to the mutant form of the region.

Vectors for telomere cloning

Vectors YAC-t1 and YAC-t2, that permit the cloning of mammalian telomeres in yeast, were constructed as illustrated in Figure 5 and as described in the materials and methods. YAC-t1 contains CEN4, ARS1, URA3 and a single copy

of TEL, the telomere repeat sequence from Tetrahymena. Upon cleavage with EcoRI and BamHI a linear molecule is generated with the TEL sequence at one end and an EcoRI cloning site at the other. In YAC-t2 the EcoRI/BamHI fragment has been replaced with a rare cutter polylinker to provide NotI, SacII, SalI, MluI, ClaI and SnaBI as additional cloning sites.

Construction of a BssHII-telomere clone library from an individual homozygous for *HD*

The steps involved in the preparation of the telomere library are summarised in Figure 6. High molecular weight genomic DNA from GUSHM1 was prepared in agarose blocks, digested to completion with BssHII and ligated to an excess of YAC-t2 which had been cleaved with BamHI and MluI and treated with alkaline phosphatase to reduce vector background. To further reduce the background of clones originating from vector alone, most of the unligated vector was then removed by electrophoresis. As a control, a fraction of the BssHII digest was ligated to MluI cleaved YAC-RC, a YAC cloning vector containing a rare cutter polylinker (Marchuk and Collins 1988). Losses incurred during vector removal were estimated by transforming fractions of this ligation from both before and after the removal of the vector, which had been carried out in parallel with that for the telomere cloning experiment. Libraries were constructed by transformation of AB1380 spheroplasts. Spheroplast transformation efficiency was tested with YCp50, giving a value of 3.5×10^5 colonies/ug vector. The ligation of BssHII digested DNA into YAC-RC gave a transformation efficiency of 5×10^3 colonies/ug insert. Therefore, we transformed ligation mix containing approximately 3 ug of insert.

sufficient to give one to two fold coverage of the clonable BssHII fragments in a standard BssHII cloning experiment.

Library screen

The 25,000 clones recovered were replicated in duplicate directly from the top agar, containing sorbitol, onto ura- plates. Filter lifts were screened with p2R88-1 and a single positive clone (Y88BT) was identified, in rough agreement with the one to two fold coverage expected. While most of the other clones recovered appear to be vector background, due to the incomplete removal of vector sequences, approximately 10% of the clones (2500 clones) were found to hybridize with a cloned Alu repeat probe. These would be expected to carry human DNA inserts, either due to the formation of (dicentric) clones with two vector arms, or the low frequency recognition of broken DNA fragments by the yeast telomerase. Up to 20% of the Alu positives that would be expected from the construction of a library in a conventional YAC vector have been recovered, suggesting minimally a five fold enrichment for telomere clones. Previously, the rate limiting step in the rapid analysis of a large number of YAC recombinants has been the difficulty in the transference of the clones embedded in top agar to a form in which they can be screened rapidly and in high density without library amplification. The method of direct replication used here has circumvented this technical problem.

Y88BT c ntains nly one YAC-t2 sequence, and carries a human telomere r peat

Chromosomes were prepared from the p2R88-1 positive clone, Y88BT, and fractionated by PFGE. The recombinant chromosome was approximately 120 kb, the size of the larger BssHII fragment detected by p2R88-1 in GUSHM1 DNA (Figure 7a). In order to determine that this clone contains the telomere of human chromosome 4p, and has not arisen from an internal BssHII fragment, Y88BT was digested with NotI and fractionated by PFGE alongside undigested DNA. NotI digestion separates the vector from the genomic DNA by cleavage within the vector polylinker. The resultant filter was hybridized sequentially with Alu, pBR322 and an oligomer, $(CCCTAA)_4$, that contains the human telomeric repeat (Figure 7b). Hybridization with Alu indicated the absence of an internal NotI site within the clone. pBR322 detected the 8 kb vector fragment only, demonstrating that all of the vector DNA had been removed from the artificial chromosome on digestion with NotI. Hybridization with $(CCCTAA)_4$ to the NotI cleaved recombinant indicates the presence of telomeric sequences on the BssHII fragment. At this stringency the oligomer did not cross-hybridize to the tetrahymena telomeric sequences on the vector. The telomeric sequences on Y88BT are human in origin and this clone therefore contains the most telomeric 120 kb of chromosome 4p.

A rare cutter restriction map of Y88BT

In order to identify and position BssHII, MluI and SalI sites within Y88BT, a restriction map was constructed using a combination of complete and partial digests (Figure 8b). This demonstrated that Y88BT starts at the BssHII site in B31 that is adjacent to an MluI site. The exact correlation between tne

position of rare cutter sites within Y88BT with that in 2R88 and B31 argues against any rearrangements in the most proximal 40 kb of the clone. The presence of the BssHII site 40 kb from the telomere indicates that the differential cleavage of this site between individuals and cell lines is a the result of a methylation polymorphism rather than the absence of a restriction site.

Genetic evidence suggests that the most likely position for the HD mutation is within Y88BT

The HD gene has been localized to the 4p16.3 band by linkage analysis with numerous DNA markers (Gusella et al. 1983; Gilliam et al. 1987a; Wasmuth et al. 1988; MacDonald et al. 1989b). However, its precise position within this small segment of the genome can only be established by a relatively small number of cases in which recombination events occur between the disease locus and the closest DNA markers. Such landmark crossovers have not provided an unequivocal placement of the HD gene, but have strongly favored a position distal to D4S90, the most telomeric marker on the genetic linkage map (MacDonald et al. 1989a; Robbins et al. 1989; Youngman et al. 1989).

Probe p88-18, from 2R88, was found to detect a rare restriction fragment length polymorphism (RFLP) when hybridised to MspI-digested human genomic DNA. The alleles, corresponding to fragments 2.0 kb and 1.8 kb, displayed frequencies of 0.98 and 0.02 respectively (N=100, where N is the number of chromosomes screened). Although not informative in the bulk of the landmark recombination events used to assign a telomeric position for *HD*, this RFLP was heterozygous in a critical meiosis from the extended Venezuela HD pedigree (Figure 9) where *HD* segregates with the rare allele.

In the mating shown, several DNA markers from 4p16.3 including D4S115, D4S111 and D4S90 have previously been ascertained to recombine with *HD* (MacDonald et al. 1989a), which if interpreted as a single recombinant, would suggest a location for *HD* proximal to these markers. However, the existence of other recombination events, incompatible with the proximal region, favors the assignment of the disease gene to the telomeric segment, distal to D4S90 (MacDonald et al. 1989a; Robbins et al. 1989). If the terminal location is correct, the MspI RFLP at D4S142 also shows recombination with *HD*, indicating that the event described above may be a double recombinant. This implied second recombination event would have to be located distal to the MspI RFLP at D4S142, as this polymorphism also shows recombination with HD. Since the polymorphic MspI site detected by p88-18 is located within the Y88BT clone, the latter would span the entire *HD* candidate region.

Discussion

The extreme telomeric location of the HD gene within 4p16.3 has made the search for closer and especially flanking markers particularly difficult. It has not been possible to identify a proven flanking marker nor to define with absolute certainty the region containing the HD mutation. However, with the exception of one event, analysis of discrete crossovers suggests a very distal location for the gene defect (MacDonald et al. 1989a; Robbins et al. 1989), telomeric to D4S90, the most distal locus previously published. Similarly, the observation of linkage disequilibrium with some but not all RFLPs has also not provided a conclusive localisation for the defect (Snell et al 1989).

We have isolated overlapping cosmids, 2R88 and B31 (D4S142) and positioned them, by pulsed field gel analysis, 200 kb distal to D4S90, within a region previously shown to be the favored location for *HD*. An RFLP at this locus, detected by p88-18, although relatively uninformative in the general population, segregates with *HD* in the Venezuela pedigree and in one sibship identifies a recombinant with the mutation. This crossover suggests a location for *HD* either proximal to D4S115 or distal to D4S142. However, owing to the existence of (minimally three) other recombination events that are incompatible with proximal location, the bulk of the recombination evidence suggests that gene may be distal to D4S142.

In order to pinpoint the position of the gene more precisely it is imperative that yet more distal and informative markers be identified. In parallel to the work described here we have used a combination of chromosome jumping (Poustka and Lehrach 1986; Poustka et al. 1987; Poustka and Lehrach 1988) and walking to isolate markers distal to this

locus. A jump from the well cut BssHII site in 2R88, in a BssHII jumping library (Poustka unpublished) and a subsequent phage walk, has allowed us to reach within 30 kb of the end of the pulsed field gel map (in preparation).

The isolation of Y88BT has provided unequivocal proof of the position of the telomere of 4p and this has for the first time set an absolute distal limit to the location of the HD gene. The telomeric candidate region of *HD* has therefore been reduced to a region spanning 100kb and the identification of the p88-18 RFLP within Y88BT indicates that the HD locus may lie within the YAC clone. Since the BssHII-telomere library was prepared from an individual homozygous for *HD*, the locus would be present in its mutant form. Therefore, we have isolated, within a single clone, the most likely of the candidate regions for the position of the HD gene.

The isolation of a YAC clone likely to carry the mutant form of the HD gene has provided the immediate potential of a functional assay in a transgenic mouse system. However, in the case of such a telomeric location, the assumption that the HD mutation directly affects a coding sequence may require reconsideration. In addition to the conventional strategy of searching for expressed or conserved sequences, based on the assumption of the existence of a cell specific "killing gene" within the region, less conventional approaches may have to be entertained, since the proximity of the defect to the telomere and the paucity of unique sequences in this region raise the specter of novel mechanisms of action of the mutation.

References

Blackburn EH (1984) Telomeres: do the ends justify the means. Cell 37:7-8

Blackburn EH, Challoner PB (1984) Identification of a telomeric DNA sequence in Trypanosoma brucei. Cell 36:447-457

Brown W (1989) Molecular cloning of human telomeres in yeast. Nature 338:774-776

Bucan M, Zimmer M, Whaley WL, Poustka A, Youngman S, Allitto BA, Ormondroyd E. et al (1989) Physical maps of 4p16.3, the area expected to contain the Huntington's Disease Mutation. Genomics (in press)

Burgers PMJ, Percival KJ (1987) Transformation of yeast spheroplasts without cell fusion. Anal Biochem 163:391-397

Burke DT, Carle GF, Olson MV, (1987) Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. Science 236:806-812

Burmeister M, Lehrach H (1989) Isolation of large DNA fragments from gels using agarase. Trends Genet 5:41

Carle GF, Olson MV (1985) An electrophoretic karyotype for yeast. Proc Natl Acad Sci USA 82:3756-3760

Cheng JF, Smith CL Cantor CR (1989) Isolation and characterisation of a human telomere. Nucl Acids Res 17:6109-6127

Chu G, Vollrath D, Davis RW (1986) Separation of large DNA molecules by contour-clamped homogeneous electric fields. Science 234:1582-1585

Church GM, Gilbert W (1984) Genomic sequencing. Proc Natl Acad Sci USA 81:1991-1995

Cross SH, Allshire RC, McKay SJ, McGill NI, Cooke HJ (1989) Cloning of human telomeres by complementation in yeast. Nature 338:771-774

Feinberg AP, Vogelstein B (1984) Addendum: a technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 137:266-267

Gilliam TC., Tanzi R, Haines JL, Bonner TI, Faryniarz AG, Hobbs WJ, MacDonald ME, et al (1987a) Localization of the Huntington's disease gene to a small segment of chromosome 4 flanked by D4S10 and the telomere. Cell 50:565-571

Gilliam TC, Bucan M, MacDonald ME,. Zimmer M, Haines JL, Cheng SV, Pohl TM et al (1987b). A DNA segment encoding two genes very tightly linked to Huntington's disease. Science 238:950-952

Greider CW, Blackburn EH (1985) Identification of a specific telomere terminal transferase activity in tetrahymena extracts. *Cell* 43:405-413

Gusella JF, Wexler NS, Conneally PM, Naylor SL, Anderson MA, Tanzi RE, Watkins PC et al (1983) A polymorphic DNA marker genetically linked to Huntington's disease. Nature 306:234-238

Herrmann BG, Barlow DP, Lehrach H (1987) A large inverted duplication allows homologous recombination between chromosomes heterozygous for the proximal t complex inversion. Cell 48:813-825

Herrmann BG, *Frischauf* A-M (1987) *Isolation of genomic DNA.* In: *Abelson* JN, Simon MI (eds) Methods in Enzymology, Vol. 152. Academic Press Inc., pp 180-183

Hieter P, Mann C, Snyder M, Davis RW (1985) Mitotic stability of yeast chromosomes: a colony color assay that measures nondisjunction and chromosome loss. Cell 40:381-392

Little RD, Porta G, Carle GF, Schlessinger D, D'Urso M (1989) Yeast artificial chromosomes with 200- to 800-kilobase inserts of human DNA containing HLA, V_K, 5S and Xq24-Xq28 sequences. Proc Natl Acad Sci USA 86:1598-1602

MacDonald ME, Anderson MA, Gilliam TC, Tranebjaerg L, Carpenter NJ, Magenis E, Hayden MR (1987) A somatic cell hybrid panel for localizing DNA segments near the Huntington's disease gene. Genomics 1:29-34

MacDonald ME, Haines JL, Zimmer M, Cheng SV, Youngman S, Whaley WL, Bucan M et al (1989a) Recombination events suggest potential sites for the Huntington's disease gene. Neuron (in press)

MacDonald ME, Cheng SV, Zimmer M, Haines JL, Poustka A, Allitto BA, Smith B et al (1989b) Clustering of multi-allele DNA markers near the Huntington's disease gene. J Clin Invest (in press)

Marchuk D, Collins FS (1988) pYAC-RC, a yeast artificial chromosome vector for cloning DNA cut with infrequently cutting restriction enzymes. Nucl Acids Res 16:7743

Martin JB, Gusella JF (1986) Huntington's disease: Pathogenesis and management. N Eng J Med 315:1267-1276

Meisson GJ, Myers RH, Mastromauro MSW, Koroshetz WJ, Klinger KW, Farrer LA, Watkins PA, (1988) Predictive testing for Huntington's disease with use of a linked DNA marker. New Eng J Med 318:535-542

Monaco AP, Bertelson CJ, Middelsworth W, Colletti CA, Aldridge J, Fischbeck KH, Bartlett R et al (1985) Detection of deletions spanning the Duchenne

muscular dystrophy locus using a tightly linked DNA segment. Nature 316:842-845

Moyzis RK, Buckingham JM, Cram LS, Dani M, Deaven LL, Jones MD, Meyne J et al (1988) A highly conserved repetitive DNA sequence, (TTAGGG)n, present at the telomeres of human chromosomes. Proc Natl Acad Sci USA 85:6622-6626

Myers RH, Leavitt J, Farrer LA, Jagadeesh J, McFarlane H, Mark RJ, Gusella JF (1989) Homozygote for Huntington's disease. Am J. Hum Genet (In press)

Pohl TM, Zimmer M, MacDonald ME, Smith B, Bucan M, Poustka A, Volinia S (1988) Construction of a NotI linking library and isolation of new markers close to the Huntington's disease gene. Nucl Acids Res 16:9185-9198

Ponzi M, Pace T, Dore E, Frontali C (1985) Identification of a telomeric DNA sequence in Plasmodium berghei. EMBO J 4:2991-2995

Poustka A, Lehrach H (1986) Jumping libraries and linking libraries: the next generation of molecular tools in mammalian genetics. Trends Genet 2:174-179.

Poustka A, Pohl TM, Barlow DP, Frischauf A-M, Lehrach H (1987) Construction and use of human chromosome jumping libraries from NotI-digested DNA. Nature 325:353-355

Poustka A, Lehrach, H (1988) Chromosome jumping: a long range cloning technique. In: Setlow JK (ed) Genetic Engineering - Principles and Methods Vol. 10 Plenum press, New York, pp 169-195

Rackwitz HR, Zehetner G, Murialdo H, Delius H, Chai JH, Poustka A, Frischauf A-M, Lehrach H (1985) Analysis of cosmids using linearization by phage lambda terminase Gene 40:259-266

Richards JE, Gilliam TC, Cole JL, Drumm ML, Wasmuth JJ, Gusella JF, Collins FS (1988) Chromosome jumping from D4S10 (G8) toward the Huntington's disease gene. Proc Natl Acad Sci 85:6437-6441

Robbins C, Theilmann J, Youngman S, Haines J, Altherr M J, Harper P S, Payne C et al (1989). Evidence from family studies that the gene causing Huntington's disease is telomeric to D4S95 and D4S90. Am J Hum Genet 44:422-425.

Sealey PG, Whittaker PA, Southern EM (1985) Removal of repeated sequences from hybridization probes. Nucl Acids Res 13:1905-1922

Shampay J, Szostak JW, Blackburn EH (1984) DNA sequences of telomeres maintained in yeast. Nature 310:154-157

Smith B, Skarecky D, Bengtsson U, Magenis RE, Carpenter N, Wasmuth JJ (1988) Isolation of DNA markers in the direction of the Huntington's disease gene from the G8 locus. Am J Hum Gene. 42:335-344

Snell RG, Lazarou L, Youngman S, Quarrell OWJ, Wasmuth JJ, Shaw DJ, Harper PS (1989) Linkage disequilibrium in Huntington's disease: an improved localisation for the gene. J Med Genet (in press)

Wasmuth JJ, Carlock LR, Smith B, Immken LL (1986) A cell hybrid and recombinant DNA library that facilitate identification of polymorphic loci in the vicinity of the Huntington disease gene. Am J Hum Genet 39:397-403

Wasmuth JJ, Hewitt J, Smith B, Allard D, Haines JL, Skarecky D, Partlow E, Hayden MR (1988) A highly polymorphic locus very tightly linked to the Huntington's disease gene. Nature 322:734-736

Wexler NS, Young AB, Tanzi RE, Travers H, Starosta-Rubinstein S, Penney JB, Snodgrass SR et al (1987) Homozygotes for Huntington's disease. Nature 326:194-197

Whaley WL, Michiels F, MacDonald ME, Romano D, Zimmer M, Smith B, Leavitt J et al (1988). Mapping of D4S98/S114/S113 confines the Huntington's defect to a reduced physical region at the telomere of chromosome 4. Nucl Acids Res 16:11769-11780

Youngman S, Shaw DJ, Gusella JF, MacDonald ME, Stanbridge EJ, Wasmuth JJ, Harper PS (1988) A DNA probe, D5 (D4S90) mapping to human chromosome 4p16.3. Nucl Acids Res 16:1648

Youngman S, Sarfarazi M, Bucan M, MacDonald ME, Smith B, Zimmer M, Gilliam TC.et al (1989) A new DNA marker [D4S90] is terminally located on the short arm of chromosome 4, close to the Huntington's disease gene. Genomics (in press)

Zehetner G, Lehrach H (1986) A computer program package for restriction map analysis and manipulation. Nucl Acids Res 14:335-349

Table 1

Probe	GM1416B					Blood				
	NotI	NruI	MluI	BssHII	SalI	NotI	NruI	MluI	BssHII	SalI
D5	850 570	350 (315) 170* (100)*	(400)* (350)* 300 80	220 170* (100)*	(290)* (270) 110	850	350 170*	1000 (570)	(240) 220 (170)* 100*	(700) 450 (400) (250)*
p2R88-1	850 280	350 (315)	300 220	(220)* 100 60 (50)*	(270) 160 (50)*	850	(570) 350	1000 (570)	(240)* (220)* 100 60 (50)*	(700) (450)* 250
88ES2.3	850 280	350 (315)	300 220	NT	90	850	350	1000	NT	250
88BH4.8	850 280	NT	300 220	(100)* (60)* (50)* 20	(270) 160 (110)*	850	(570) 350	NT	(50)* 20	(450)* 250

SUBSTITUTE SHEET

Table 1 (cont)

Probe	GM1416B					Blood				
	N/R	N/M	N/B	N/S	R/M	N/R	N/M	N/B	N/S	R/M
D5	350 (315) 170* (100)* 35	300 80 20	220 170* (100)* 60	(290)* (270) 110 80	300 170* 100* 80	350 170*	850 (570)	(240) 220 (170)* 100*	(700) 450 (400) 250*	(570) (400) 350 170*
p2R88-1	350 (315) 280	300 220	(200)* 100 60 (50)*	(270) 160	300 220	350	850 570	(240)* (220)* 100 60 (50)*	(700) (450)* 250	(570) 350
88ES2.3	NT	NT	NT	90	NT	NT	NT	NT	250	NT
88BH4.8	NT	NT	(100)* (60)* (50)* 20	NT	NT	NT	NT	(50)* 20	NT	NT

SUBSTITUTE SHEET

Table 1 (cont)

Probe	GM1416B					Blood				
	R/B	R/S	M/B	M/S	B/S	R/B	R/S	M/B	M/S	B/S
D5	220 170* (100)*	(260) 100 (60)	180 170* (100)* 80	(290)* (270) (110) 50	(220) 170* 90	(240) 220 (170)* 100*	(160) 100	(240) 220 (170)* 100*	450 (400) 250*	(240) (220) (170)* 100* 90
p2R88-1	(220)* 100 60 (50)*	(260) 160	100 60 (50)*	160 130 (50)*	(50)* 11	(240)* (220)* 100 60 (50)*	(350) 250 (160)* 100*	(240)* 220* 100 60 (50)*	(700) 450 250	100 60 (50)*
88ES2.3	NT	90	NT	90	NT	NT	250	NT	250	NT
88BH4.8	(100)* (60)* (50)* 20	NT	(100)* (60)* (50)* 20	NT	(50)* 20	(50)* 20	NT	(50)* 20	NT	(50)* 20

SUBSTITUTE SHEET

Table 2

Probe	Size of partial bands (kb)
pBR322	23 73 83 123
p2R88-1	50 60 73 83 100 123

Legends to tabl s

Table 1

Summary of rare cutter restriction fragments (kb; +/- 10%) detected by D5 and by probes isolated from 2R88.

Weakly detected bands, generated by partial digestion of a restriction site, are indicated in parentheses. An asterisk indicates the presence of a weak band detected by cross-hybridization to another locus. p2R88-1 and 88BH4.8 both crosshybridize to SalI and BssHII fragments of approximately 50 kb, and D5 to BssHII and NruI fragments of approximately 170 kb that are located proximal to this map but within 700 kb of the distal end of the map. p2R88-1 also crosshybridizes with a locus on the 220 kb BssHII fragment detected by D5, and 88BH4.8 to a locus that detects the same BssHII fragments as p2R88-1.

N=NotI, M=MluI, R=NruI, S=SalI and B=BssHII. NT = not tested.

Table 2

Sizes of partial BssHII restriction fragments (+/- 10%) of Y88BT, detected by hybridization with pBR322 and p2R88-1 (Figure 8a).

Legends to figures

Figure 1: Characterization of the 2R88 and B31 cosmids

a) Restriction maps of the 2R88 and B31 cosmids indicating the position of the rare cutter restriction sites BssHII MluI and SalI. Fragments that were used as probes against Southern filters are outlined above the map. These

fragments that flank the BssHII sites, 88E1.8 is the 1.8 kb EcoRI fragment containing the SalI site and 88ES2.3 the 2.3 kb EcoRI/SalI fragment (the SalI site is in the vector). p2R88-1 is a Sau3A/PstI fragment of 750 bp in pGEM4 that originates from the 2.2 kb BamHI/HindIII fragment and p88-18 contains a Sau3A fragment in pGEM4 that maps to the 700 bp EcoRI/HindIII fragment.

b) Demonstration that the BssHII and SalI sites within the 2R88 cosmid are unmethylated in genomic DNA isolated from the cell line GM1416B. 88BH2.0 and 88E1.8 were hybridized in turn to a filter containing GM1416B DNA that had been digested with HindIII only and with HindIII and a rare cutter retriction enzyme as indicated. The 6.8 kb HindIII fragment detected by 88BH2.0 is reduced to 2.0 and 5.2 kb by BssHII and ClaI respectively. Fainter bands are indicative of cross-hybridization to other loci. The 4.5 kb HindIII fragment detected by 88E1.8 is reduced to 2.8 kb by SalI cleavage.

Figure 2: Pulsed field gel analysis of D5 and p2R88-1
Hybridization of the probes p2R88-1 and D5 to a pulsed field filter containing both GM1416B and leukocyte DNA that has been digested with rare cutting restriction enzymes as indicated. Electrophoresis was for 36 hrs at 5 V/cm in 0.75% agarose/0.25 x TBE with a pulse time of 50 sec. LM indicates the region of limiting mobility.

N=NotI, M=MluI R=NruI S=SalI and B=BssHII.

Figure 3

A long range restriction map spanning 350 kb and illustrating the physical linkage between the D5 (D4S90) and 2R88/B31 (D4S142) loci and the position of the 4p telomere (indicated as restriction sites in italics). Linkage between D5 and 2R88 is demonstrated by hybridization to the same 350 kb NruI, 1000 kb MluI (Table 1), and 850 kb NotI (Table 1) fragments in leukocyte DNA and to the same 350 kb NruI fragment and partial 300 kb MluI and 850 kb NotI (Table 1) fragments in DNA from GM1416B cells. The SalI fragments detected by D5 in DNA from leukocytes and GM1416B cells are 450 and 100 kb respectively, and the SalI sites that flank this fragment in GM1416B DNA can be positioned by double digestion with NotI, MluI and NruI. This places a SalI site 250 kb from the distal end of the map which is the size of the SalI fragment detected by p2R88-1 in leukocyte DNA. It is also the sum of the 160 and 90 kb SalI fragments that hybridize with p2R88-1 and 88ES2.3 respectively (probes that flank the well cleaved SalI site in the 2R88 cosmid that is not cut in leukocyte DNA). This position of 2R88 is confirmed by the 130 kb SalI/MluI double digestion product detected by p2R88-1 in GM1416B DNA. The orientation of the cosmid is indicated by the hybridization of p2R88-1 and D5 to GM1416B DNA that has been digested with SalI and BssHII, positioning the BssHII sites flanking the D5 BssHII fragment. Consideration of these data, alongside the predetermined position of the SalI site within 2R88, allows only one orientation of the 2R88 cosmid. Parentheses indicate that a site is very partially cleaved. The NotI* and MluI* sites are available for restriction on one chromosome only and situated on opposite chromosomes. It was not possible to be certain that all of the BssHII sites within this region had been positioned.

Figure 4: BssHII polymorphisms detected by p2R88-1.
Hybridization of p2R88-1 to a Southern blot of BssHII digested DNA from a number of individuals. PFGE was in 0.9% agarose with a pulse time of 40 sec for 40 hrs. The DNA samples are as follows: lane 1: GUSHM1 ; lane 2: GUSHM3; lane 3: GUS641; lane 4; GUS 115. GUSHM1and GUSHM3 are individuals from the Venezuela pedigree likely to be homozygous for *HD*. GUS641 and GUS115 are individuals from the Venezeula pedigree not carrying the HD mutation

Figure 5: Construction of the telomere cloning vectors YAC-t1 and YAC-t2.
YAC-t1 was generated by the replacement of the BamHI/SalI fragment of YCp50 with the BamHI/XhoI fragment from YAC4 containing the Tetrahymena telomeric repeat. Cleavage with EcoRI and BamHI creates a linear molecule flanked by a Tetrahymena telomere and an EcoRI cloning site. YAC-t2 was constructed by the replacement of the BamHI/EcoRI fragment from YAC-t1 with a rare cutter polylinker providing additional cloning sites as indicated.

Figure 6: Construction of the BssHII telomere library
High molecular weight DNA from GUSHM1 was digested to completion with BssHII and ligated into the MluI site of MluI/BamHI cut and phosphatased YAC-t2. Recombinant molecules were transformed into AB1380 spheroplasts and the telomere of human chromosome 4p was subsequently identified by hybridization to p2R88-1.

Figure 7: Demonstration that the structure of Y88BT is that of a telomere clone

a) Y88BT was digested with NotI and fractionated alongside the undigested clone (UD). YP148 chromosomes were used as a size marker. Electrophoresis was in a 0.9% agarose gel with a pulse time of 40 sec for 36 hrs to give separation up to approximately 500 kb.

b) Hybridization to the resultant filter from the gel in a) with Alu, pBR322 and the oligomer $(CCCTAA)_4$.

Figure 8: A rare cutter restriction map of Y88BT

a) Hybridisation of pBR322 and p2R88-1 to a Southern blot of BssHII partial digests of Y88BT. Agarose blocks containing approximately 1 ug of Y88BT DNA were digested with decreasing amounts of enzyme (as indicated) in 200 ul for 3 hrs. PFGE was in 1% agarose with a pulse time of 5 sec for 24 hrs, to provide a gel resolution of 100-150 kb. Lambda multimers and lambda DNA digested with HindIII were used as size markers. LM is the region of limiting mobility.

b) A rare cutter restriction map of the Y88BT clone with the enzymes MluI, SalI, and BssHII (NotI and SalI cut within the polylinker). The map was constructed from a combination of data generated by PFGE analysis of complete and partial digests using the conditions described in a). The resultant filters were hybridized sequentially with p2R88-1, Alu, the human telomeric repeat oligomer, $(CCCTAA)_4$, and with pBR322. The SalI and MluI sites were positioned with information gained from complete digestion of the YAC. The number and relative positions of the BssHII sites were determined from the fragments observed by hybridization with pBR322

and p2R88-1 to the partial digests as shown in a). The distances between these sites (estimated to be +/- 10%) were judged from the size of complete digest bands detected by Alu, and those close to the telomere from the hybridization of $(CCCTAA)_4$ to the filter in a).

Figure 9: Presence of D4S142 in Y88BT and segregation of the locus in an HD recombinant family

a) p88-18 probe was hybridized to blots containing 5 ug of the following DNAs digested with MspI: lane 1: HHW842, a human-hamster somatic cell hybrid containing human chromosome 5 and an interstitial deletion human chromosome 4 which retains the terminal portion of 4p16.3; lane 2: hamster cell line tsH1 spiked with 100ng of DNA from yeast containing Y88BT; lane 3: tsH1; lane 4: HHW847, a human hamster hybrid containing several human chromosomes including human chromosome 5 and a t(4;21) chromosome in which 4p16.2-pter is absent; lane 5: the father (F) of the nuclear family shown in b); lane 6: the mother (M); lane 7: an affected offspring showing recombination between *HD* and the D4S142 MspI RFLP; and lane 8: an affected offspring (NR) showing no recombination. The presence of the expected allelic fragment in HHW842 but not in HHW847 establishes that the locus maps to the terminal portion of 4p16.3. The fragment representing allele 2 is seen in Y88BT since this allele segregates with *HD* in the extended Venezuelan pedigree from which the homozygous *HD* cell line, used for library construction, was derived. The results in lanes 5-8 can be interpreted by referring to b).

b) The nuclear family shown derives from the Venezuela HD kindred and consists of an affected father (F), unaffected mother (M) and two affected

progeny (NR and R) with differing genotypes for several 4p16.3 markers as previously reported (MacDonald et al, 1989a). The sex of the children is not given to protect the confidentiality of the family and to preserve the blinded status of our clinical collaborators. The normal phase of the various marker alleles with respect to the HD chromosome, shown in individuals F and NR and symbolized by the filled chromosome schematic, is known from other affected sibs and close relatives in the remainder of the Venezuela pedigree. The affected individual designated as R has been reported to display recombination between the markers D4S115, D4S111 and D4S90 and the disease gene. The data shown in a) indicates D4S142 also displays recombination, leaving two possible locations for the HD gene: 1) the region between D4S10 and D4S115 and 2) the region between D4S142 and the telomere in the case of a double recombination event. The latter location is favored, however, because of several independent crossovers in other families not informative for the D4S142 RFLP that are inconsistent with the HD gene being proximal to D4S111 and D4S90.

CLAIMS

1. YAC - clone Y88 BT (NCYC 2336).

2. A nucleic acid probe capable of hybridising with YAC - clone Y88 BT (NCYC2336).

3. A nucleic acid fragment capable of hybridising with YAC - clone Y88 BT (NCYC2336) and containing a coding sequence.

4. A nucleic acid fragment according to claim 3 containing at least a portion of the coding sequence of a wild type or mutant gene which, in mutant form, is responsible for Huntington's Disease.

5. An expression vector comprising a nucleic acid fragment according to claim 3 or claim 4 in expressible form.

6. A virus particle containing a nucleic acid fragment according to claim 4 in expressible form as part of the genomic nucleic acid of the virus.

7. A cell transfected with a vector according to claim 5 or infected with a virus particle according to claim 6.

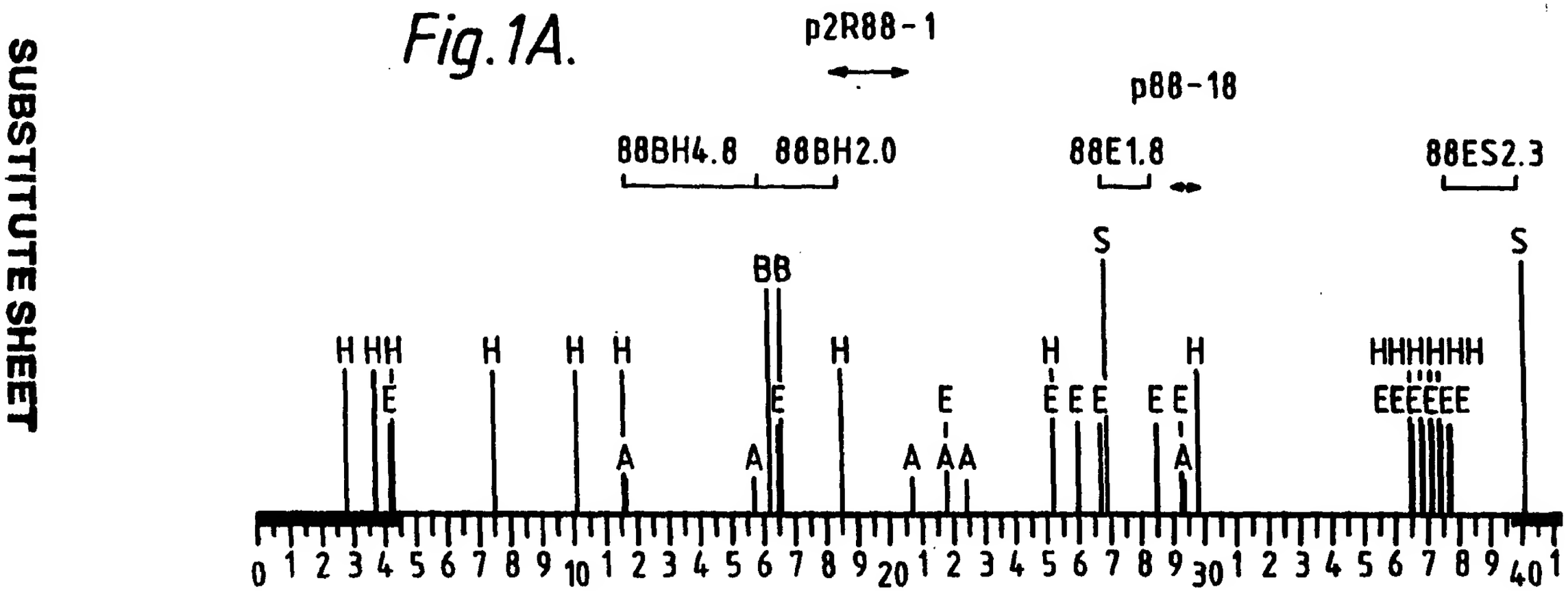
8. A process for producing a polypeptide comprising culturing a cell according to claim 7 under conditions permitting expression of the coding sequence.

9. A polypeptide encoded by a coding sequence of YAC - clone Y88 BT (NCYC2336) or a fragment thereof.

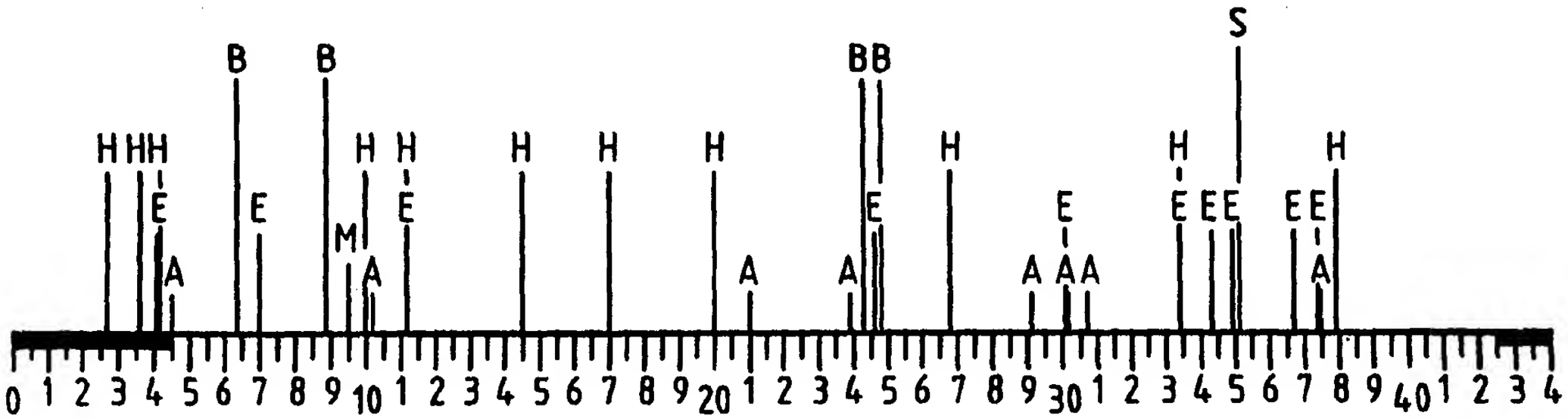
10. An antibody against a polypeptide according to claim 9 or produced by a process according to claim 8.

11. Use of a nucleic acid probe, fragment, expression vector, virus particle, cell, polypeptide or antibody according to any one of claims 1 to 7, 9 and 10 in diagnosis or therapy of Huntington's Disease.

Fig.1A.



MAP of 2R88 COSMID (Total length = 41.24kb)

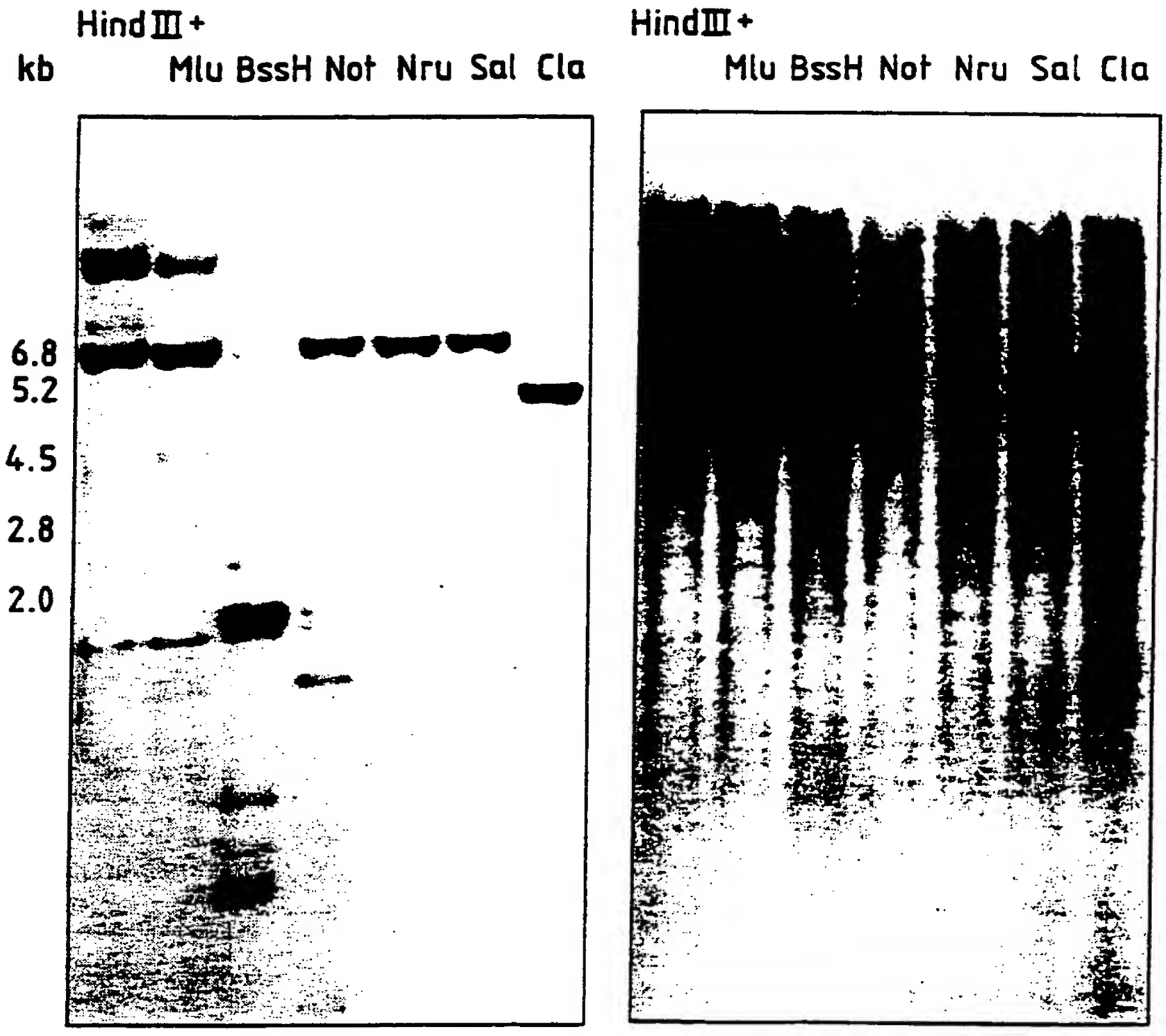


MAP of b31 COSMID (Total length = 44.01kb)

BssHII B, HindIII H, EcoRI E, MluI M, BamHI A, SalI S

PCT/GB90/01481

Fig. 1B.


Hind III +
kb
Mlu BssH Not Nru Sal Cla
6.8
5.2
4.5
2.8
2.0
88BH2.0
Hind III +
Mlu BssH Not Nru Sal Cla
88E1.8

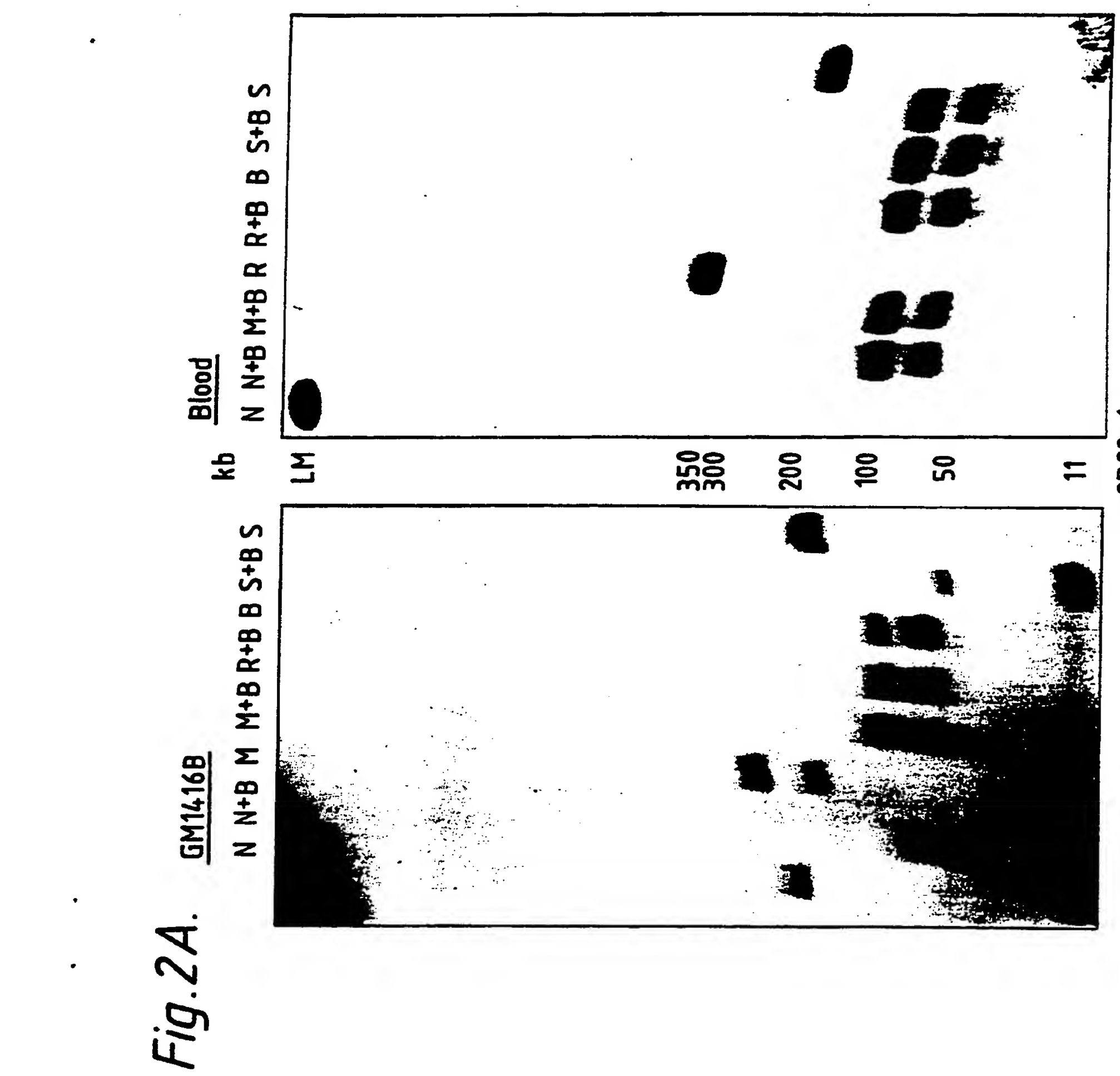

A.
GM1416B
N N+B M M+B R+B B S+B S
kb
LM
350
300
200
100
50
11
Blood
N N+B M+B R R+B B S+B S
p2R88-1

Fig.2

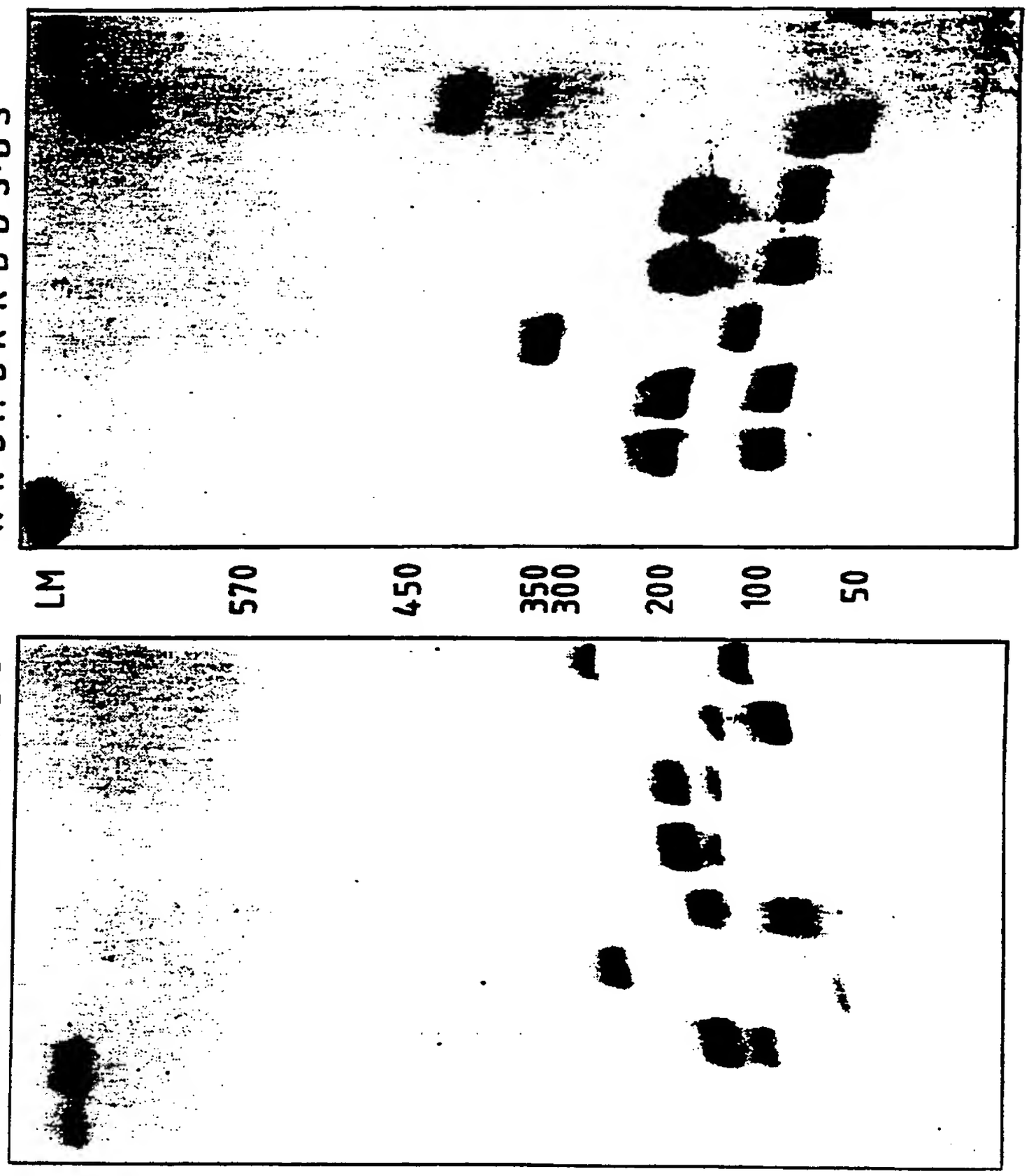
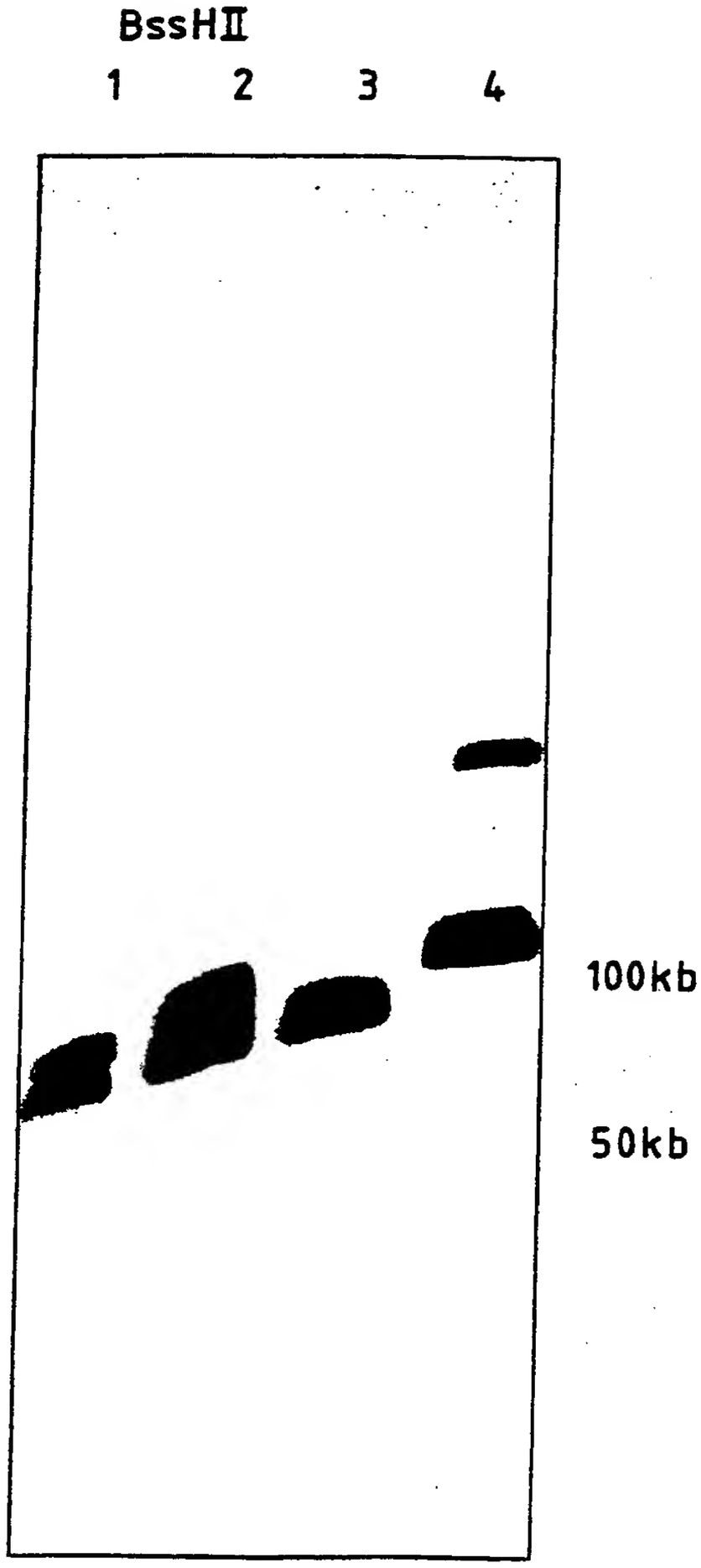

GM1416B
N N+B M M+B R+B B S+B S
Blood
N N+B M+B R R+B B S+B S
kb
LM
570
450
350
300
200
100
50
DS

Fig.2B.

Fig. 4.



p2R88-1

Fig. 5.

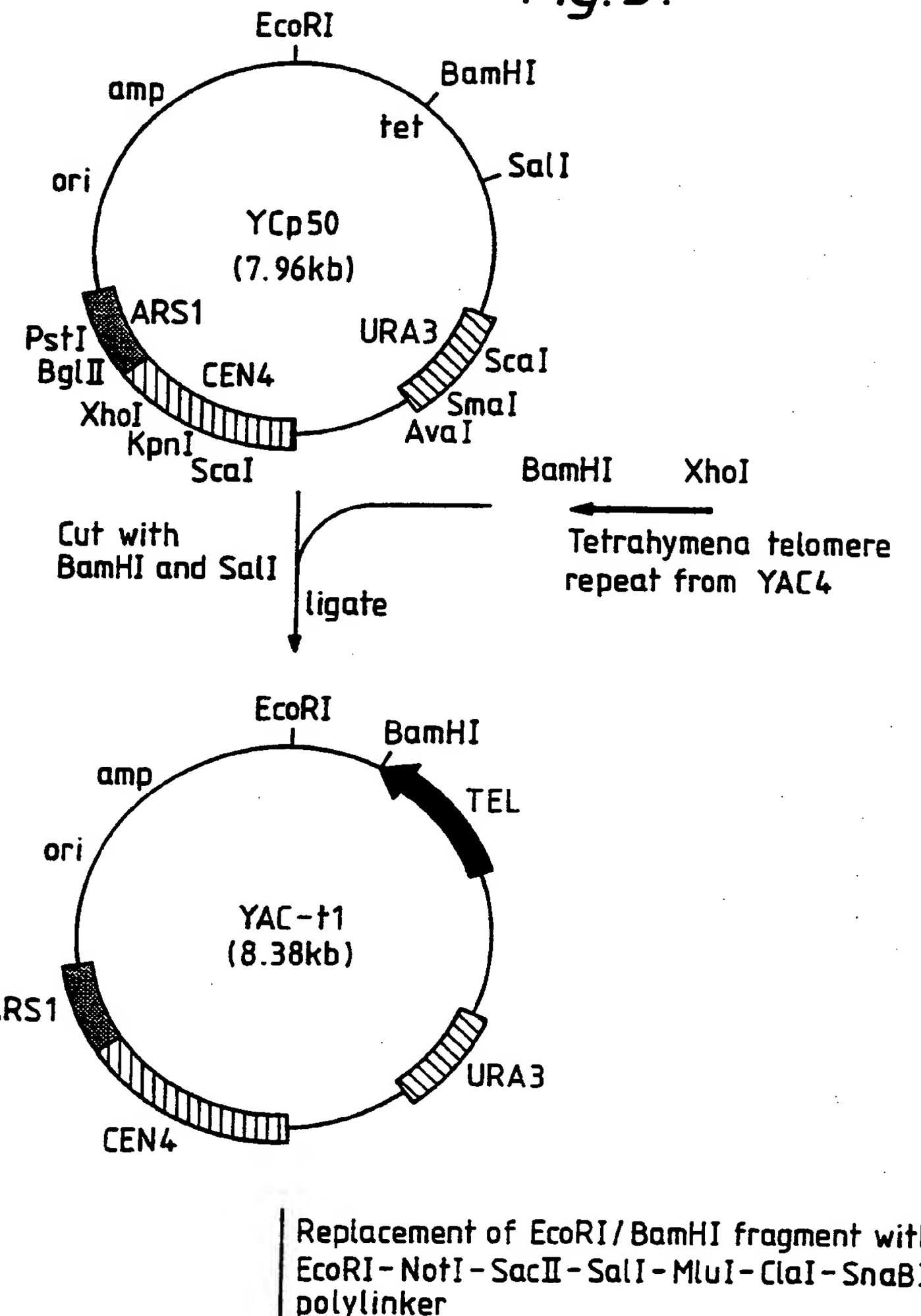

EcoRI
BamHI
amp
tet
SalI
ori
YCp50
(7.96kb)
ARS1
PstI
BglII
CEN4
URA3
ScaI
SmaI
XhoI
AvaI
KpnI
ScaI
BamHI
XhoI
Cut with
BamHI and SalI
Tetrahymena telomere
repeat from YAC4
ligate
EcoRI
BamHI
amp
TEL
ori
YAC-t1
(8.38kb)
ARS1
URA3
CEN4
Replacement of EcoRI/BamHI fragment with
EcoRI-NotI-SacII-SalI-MluI-ClaI-SnaBI-BamHI
polylinker
YAC-t2
(8.05kb)

Fig.6.

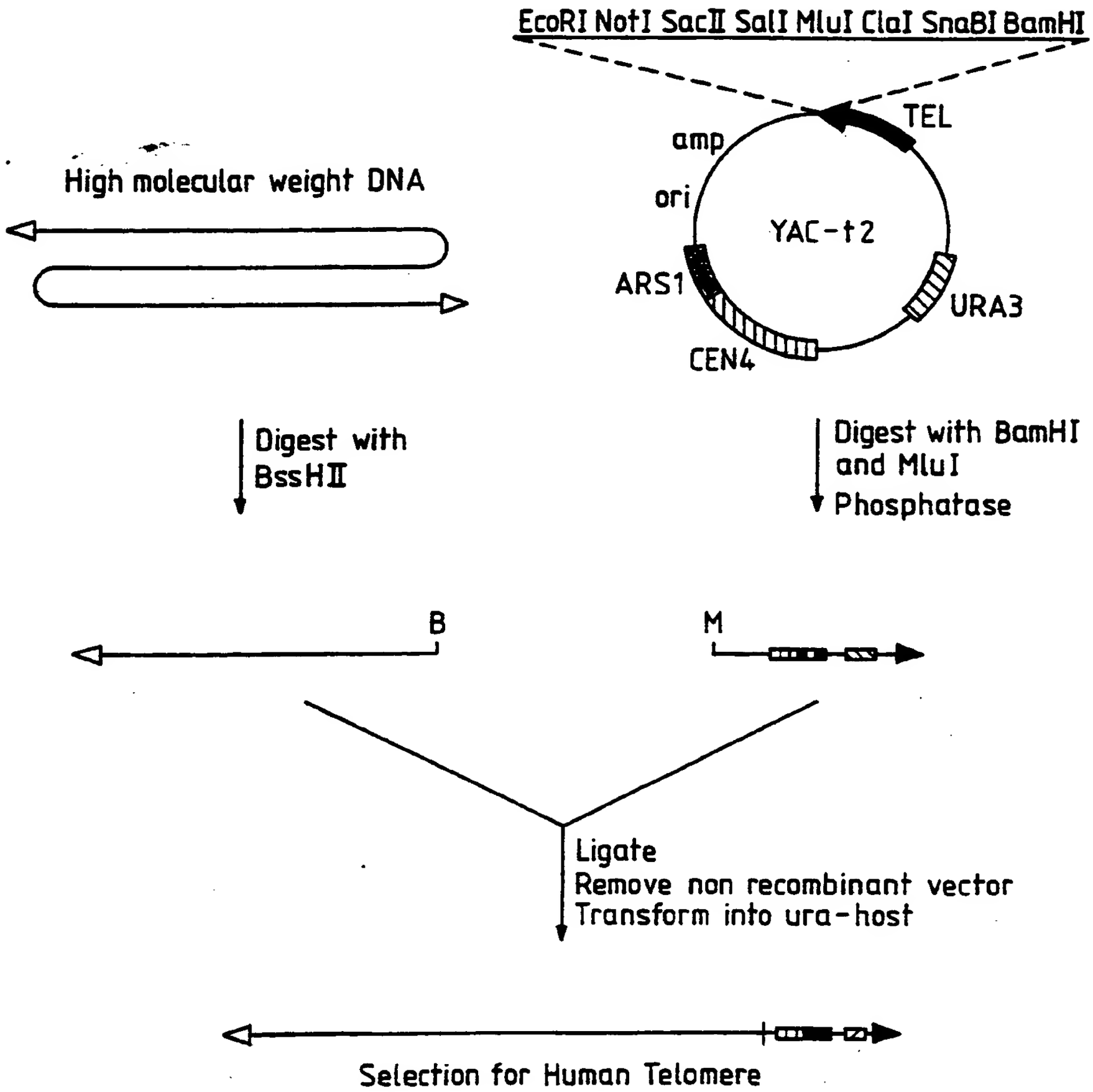

EcoRI NotI SacII SalI MluI ClaI SnaBI BamHI
TEL
amp
ori
YAC-t2
ARS1
URA3
CEN4
High molecular weight DNA
Digest with BssHII
Digest with BamHI and MluI
Phosphatase
B
M
Ligate
Remove non recombinant vector
Transform into ura-host
Selection for Human Telomere

Fig. 7A.

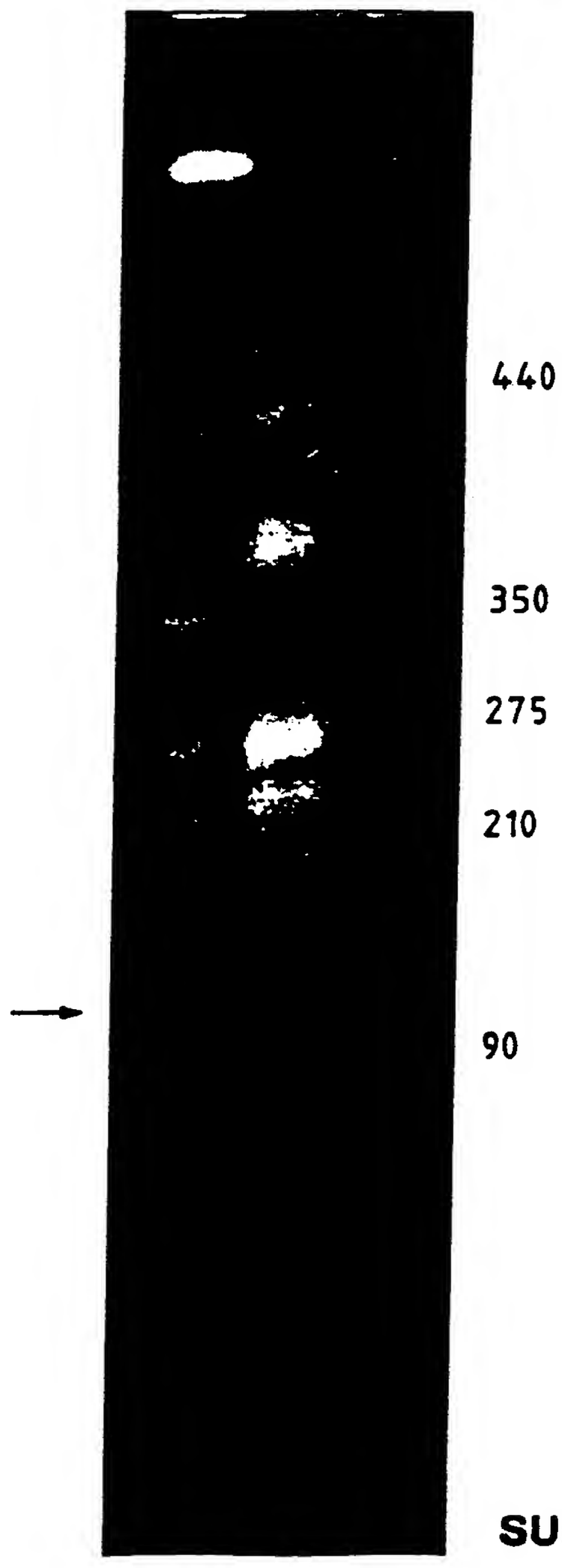


Fig. 7B.

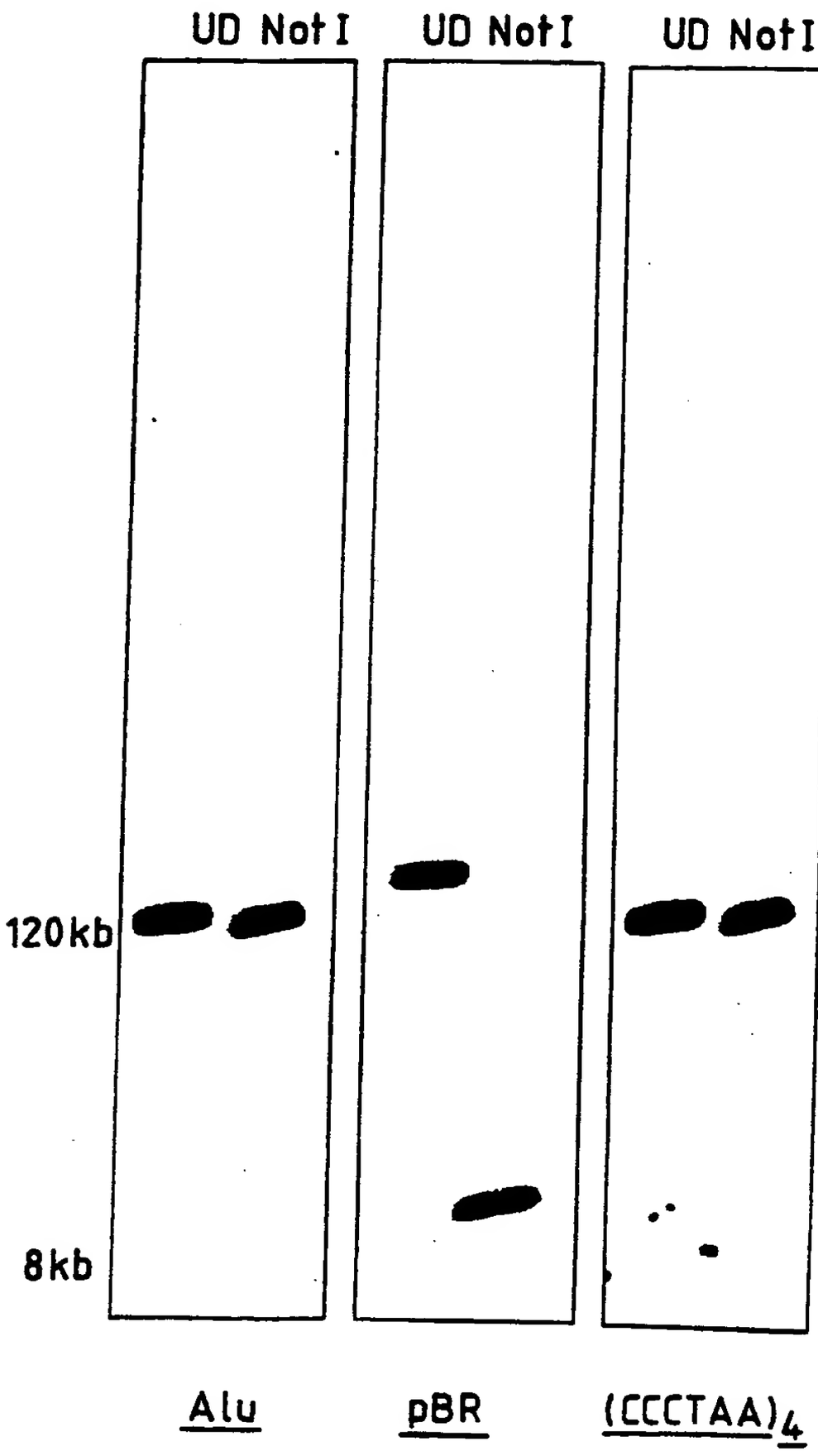


Fig.8A.

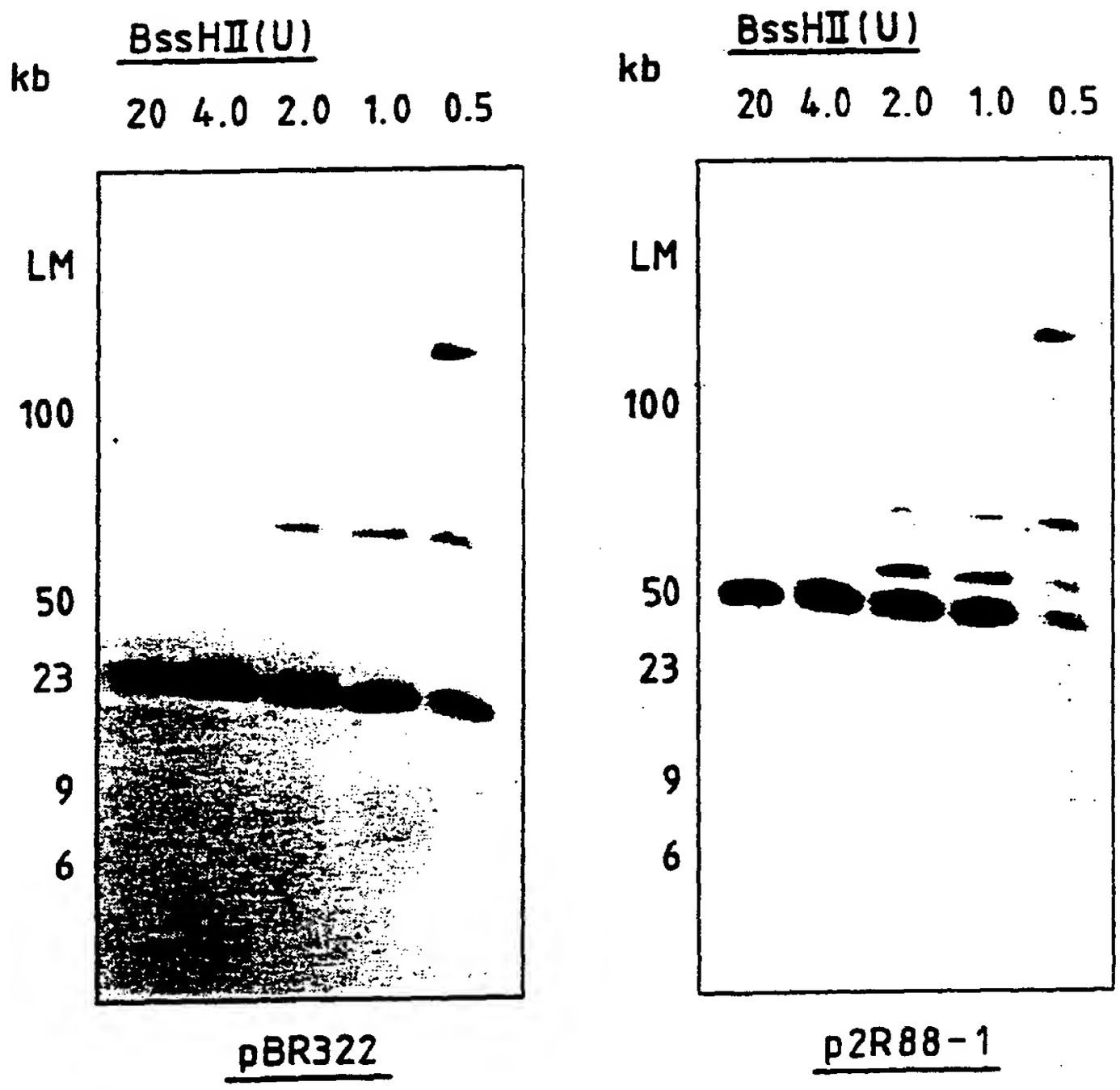
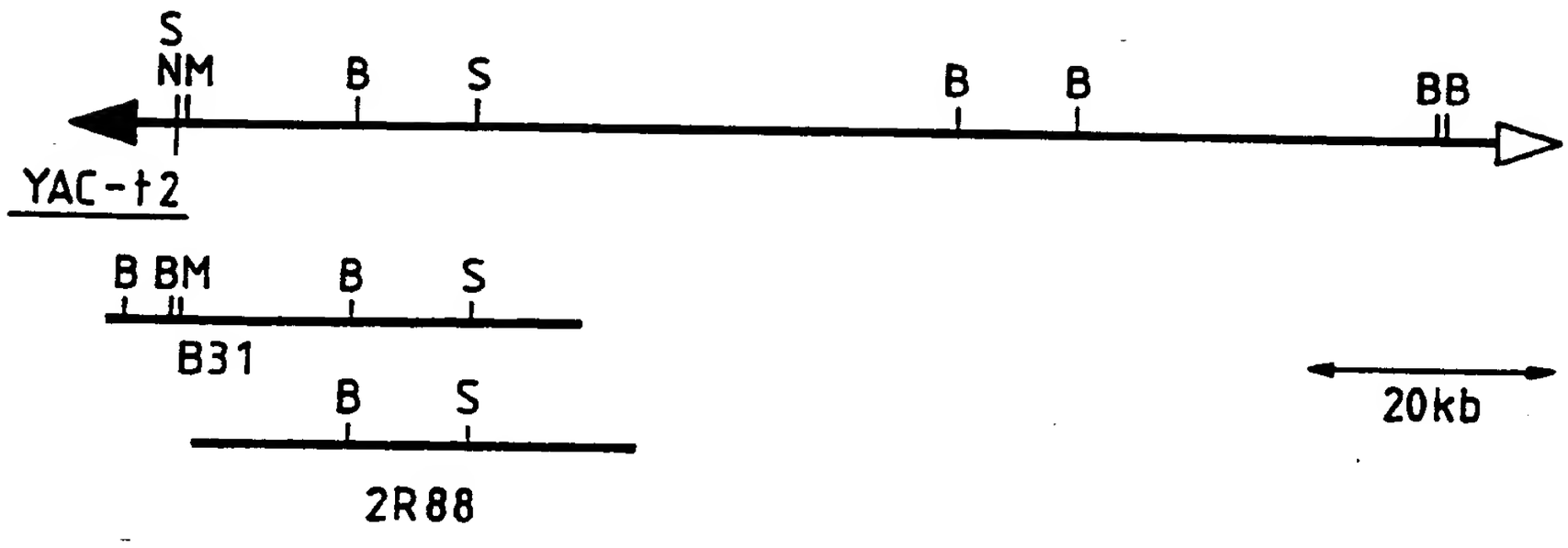

BssHII(U)
kb
20 4.0 2.0 1.0 0.5
LM
100
50
23
9
6
pBR322
BssHII(U)
kb
20 4.0 2.0 1.0 0.5
LM
100
50
23
9
6
p2R88-1

Fig. 8B.



N=NotI M=MluI S=SalI B=BssHII

Fig. 9A.

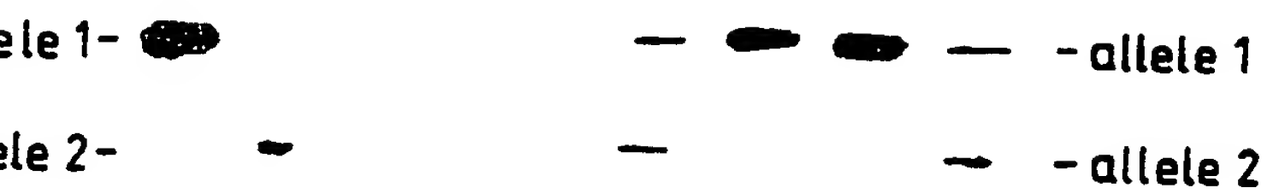


Fig. 9B.

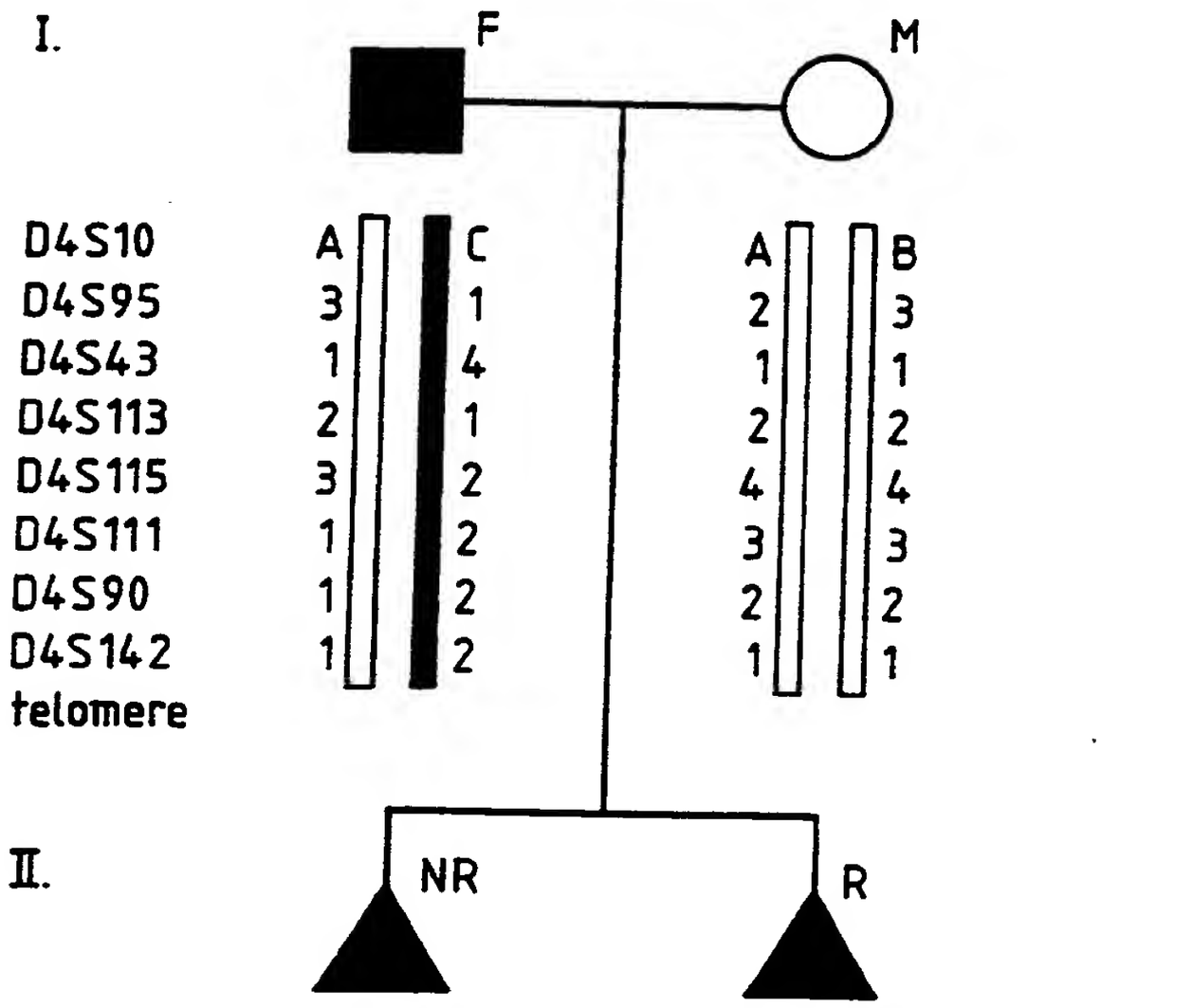
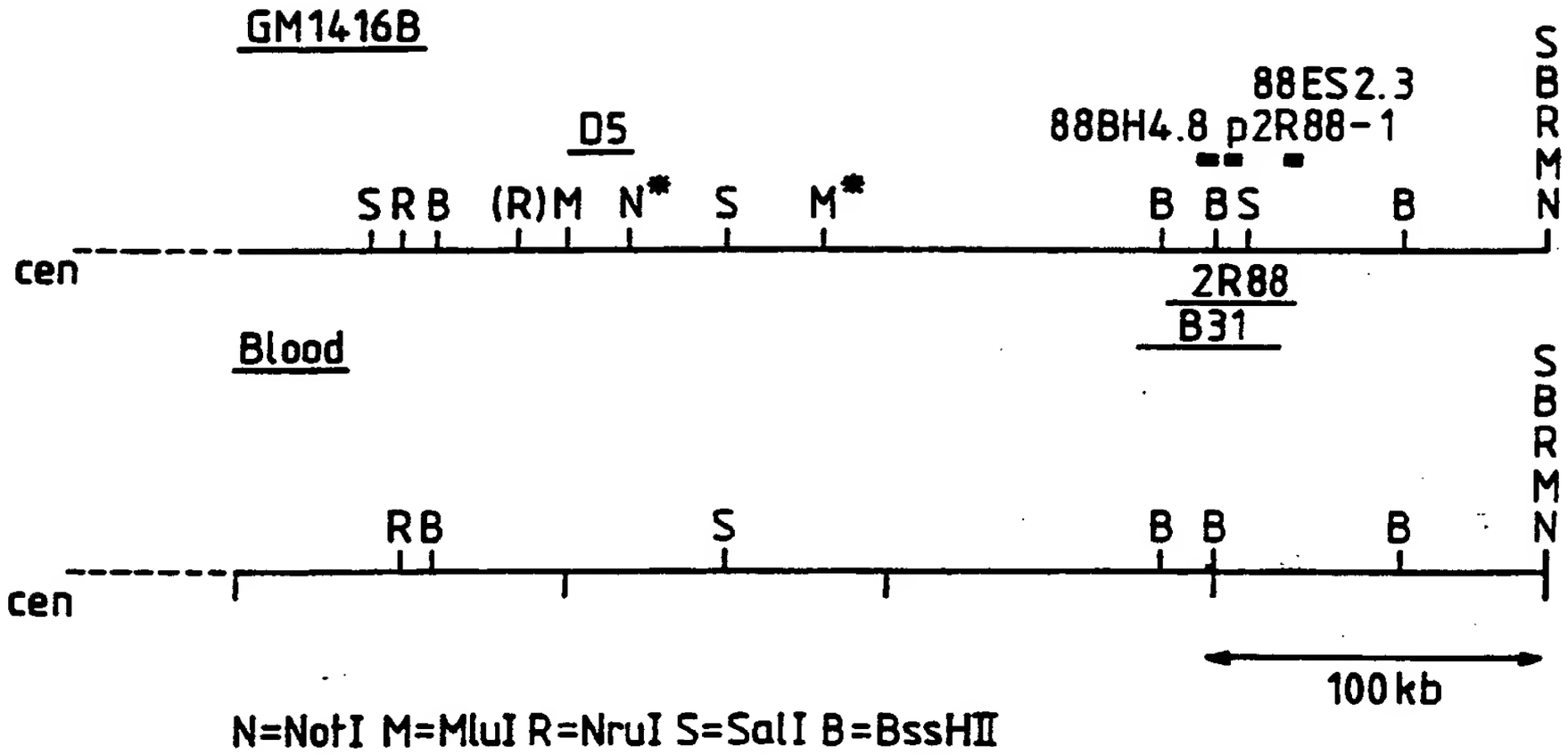


Fig. 10.


GM1416B
D5
88ES2.3
88BH4.8 p2R88-1
S B R M N
S R B (R) M N* S M*
B B S B
cen
2R88
B31
Blood
S B R M N
R B S B B B
cen
100 kb
N=NotI M=MluI R=NruI S=SalI B=BssHII

INTERNATIONAL SEARCH REPORT

International Application No PCT/GB 90/01481

I. CLASSIFICATION OF SUBJECT MATTER (If several classification symbols apply, indicate all)[6]

According to International Patent Classification (IPC) or to both National Classification and IPC

IPC5: C 12 N 15/12, C 12 Q 1/68, C 12 Q 1/02, G 01 N 33/53

II. FIELDS SEARCHED

Minimum Documentation Searched[7]

Classification System	Classification Symbols
IPC5	C 12 N; C 12 Q

Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in Fields Searched[8]

III. DOCUMENTS CONSIDERED TO BE RELEVANT[9]

Category *	Citation of Document,[11] with indication, where appropriate, of the relevant passages[12]	Relevant to Claim No.[13]
P,X	Dialog Information Services, File 154, Medline 83-90, NLM accession no. 90196008, Bates GP: "A yeast artificial chromosome telomere clone spanning a possible location of the Huntington disease gene", Am J Hum Genet (UNITED STATES) APR 1990, 46 (4) P762-75	1-11
	--	
X	US, A, 4666828 (JAMES F. GUSELLA) 19 May 1987, see the whole document	1-11
	--	
X	Dialog Information Services, File 154, Medline 83-90, NLM acession no. 89098324, Whaley WL: "Mapping of D4S98/S114/S113 confines the Huntington's defect to a reduced physical region at the telomere of chromosome 4", Nucleic Acids Res (ENGLAND) Dec 23 1988, 16 (24) P 11769-80	1-11
	--	

* Special categories of cited documents: [10]

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance, the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance, the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report
21st December 1990	25 JAN 1991
International Searching Authority	**Signature of Authorized Officer**
EUROPEAN PATENT OFFICE	MISS D. [illegible]ALCZYK

Form PCT/ISA/210 (second sheet) (January 1985)

66

ANNEX TO THE INTERNATIONAL SEARCH REPORT ON INTERNATIONAL PATENT APPLICATION NO.PCT/GB 90/01481

SA 40633

This annex lists the patent family members relating to the patent documents cited in the above-mentioned international search report.
The members are as contained in the European Patent ffice EDP file on 28/11/90
The European Patent office is in no way liable for theseparticulars which are merely given for the purpose of information.

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US-A- 4666828	19/05/87	NONE	

For more details about this annex : see Official Journal of the European patent Office, No. 12/82

EPO FORM P0479